

UNIVERSIDADE ESTADUAL DE CAMPINAS

UNICAMP

INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO

" O MODELO DE REGRESSÃO LINEAR BI-SEGMENTADO NA  
ESTIMAÇÃO DO LIMAR DE ANAEROBIOSE "

ALONSO MAZINI SOLER

CAMPINAS (SP)

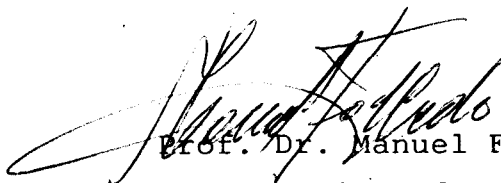
1988

UNICAMP  
BIBLIOTECA CENTRAL

"O MODELO DE REGRESSÃO LINEAR BI-SEGMENTADO NA ESTIMAÇÃO  
DO LIMÍAR DE ANAEROBIOSE"

Este exemplar corresponde a  
redação final da tese devi-  
damente corrigida e defendi-  
da pelo Sr. Alonso Mazini  
Soler e aprovada pela Comis-  
são julgadora.

Campinas, 04 de julho de 1988.



Prof. Dr. Manuel Folledo  
orientador

Dissertação apresentada ao  
Instituto de Matemática ,  
Estatística e Ciência da  
Computação, IMECC/UNICAMP,  
como requisito parcial pa-  
ra obtenção do título de  
Mestre em Estatística.

Orientador:

Prof. Dr. Manoel Folledo

As informações observacionais utilizadas neste trabalho foram coletadas através de um protocolo experimental desenvolvido pelo Laboratório de Hemodinâmica e Função Pulmonar do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto.

Aos meus pais, Alonso e Celina.

Ao João e à Camila.

À minha esposa Júlia:

## AGRADECIMENTOS

Ao professor Manoel Folledo pela amizade e pela orientação;

Aos amigos do IMECC-UNICAMP, em especial aos colegas de curso e aos professores que contribuíram para a minha formação;

Aos amigos do Departamento de Matemática do IPEA-UNESP pela confiança que depositaram em meu trabalho;

À Maria Helena pela grande amizade;

Aos amigos Dr. Gallo e Barreto pela colaboração e incentivo;

Às instituições CNPQ, CAPES e FAPESP pelo apoio financeiro recebido;

## NOTAÇÃO

- .  $A$   $\equiv$  Parâmetro a ser estimado;
- .  $\hat{A}$   $\equiv$  Estimativa de mínimos quadrados ordinários;
- .  $\hat{A}^+$   $\equiv$  Estimativa de mínimos quadrados ordinários sujeita a restrições;
- .  $\hat{A}^*$   $\equiv$  Estimativa de mínimos quadrados generalizados;
- .  $\hat{A}^\#$   $\equiv$  Estimativa de mínimos quadrados generalizados sujeita a restrições.

CAPÍTULO	V. O Modelo Estatístico Na Resolução Do Problema	
	Biológico	
	V.1. Introdução .....	51
	V.2. Formulação Do problema .....	56
	V.2.1. Procedimento De Estimação Do Ponto De	
	Junção .....	61
	V.3. Ilustração - Um Estudo De Simulação .....	64
	V.3.1. Discussões .....	68
CAPÍTULO	VI. Aplicação	
	VI.1. O Delineamento Experimental .....	78
	VI.2. Resultados .....	82
	VI.3. Discussões .....	87
SUMÁRIO .....		98
SUMMARY .....		99
APÊNDICE 1 .....		100
APÊNDICE 2 .....		118
APÊNDICE 3 .....		122
REFERÊNCIAS BIBLIOGRÁFICAS .....		126



## CAPÍTULO I

### INTRODUÇÃO

Suponha que um sistema esteja sujeito a determinadas modificações instantâneas relevantes, previsíveis ou não, em suas características estocásticas, capazes de influir na manutenção contínua de seu perfeito funcionamento.

Não seria difícil enumerar uns tantos exemplos de fenômenos que poderiam ser caracterizados desta maneira, nem tampouco seria estafante avaliar os transtornos decorrentes desta "instabilidade" estocástica, portanto, não se requer estranheza deparar com um grande número de trabalhos científicos dedicados à análise de problemas correlatos. Como ilustração, pode-se citar determinadas situações, provenientes dos trabalhos de Controle Estatístico de Qualidade, nas quais, determinadas modificações no comportamento probabilístico dos "defeitos" do processo de produção, influenciam decisivamente na preservação dos padrões de qualidade adotados.

De interesse capital nos estudos dedicados à análise deste problema, destacam-se o instante e a forma como se processa tal transição probabilística. Assim, retira-se da literatura específica, tanto trabalhos relativos a casos nos quais o momento da transição se dá de forma abrupta e repentina quanto casos nos quais se observa uma transição suave e quase imperceptível.

De qualquer maneira como se analise o momento da transição, este cumpre o papel de "divisor de águas", particionando sequencialmente (temporalmente) as possíveis respostas obtidas de um sistema em dois ou mais conjuntos heterogêneos entre si sob o ponto de vista estocástico. Bastante citado na literatura estatística, o problema é reconhecido através da denominação "Change Point Problem".

Para efeito deste trabalho, restringe-se a gama de ocorrências de possíveis alterações probabilísticas do sistema a apenas uma, isto é, as respostas obtidas de um sistema passam de um para outro estado probabilístico, num dado instante e de uma tal forma.

Formalizando matematicamente as idéias expostas até então, suponha uma sequência ordenada (temporal) de variáveis aleatórias  $X_1, X_2, \dots, X_N$ . Diz-se que ocorreu um "change point" em  $\tau$ , onde  $\tau$  é inteiro no intervalo  $[2, N-1]$ , caso:

$$\begin{aligned} X_i &\sim F_1(X, \theta_1), & i = 1, 2, \dots, \tau \\ X_i &\sim F_2(X, \theta_2), & i = \tau+1, \tau+2, \dots, N \end{aligned}$$

onde,  $F_1(\dots)$  e  $F_2(\dots)$  supõe-se possuírem formas funcionais comuns e conhecidas, diferindo apenas pelos valores desconhecidos dos parâmetros  $\theta_1 = (\theta_{11}, \theta_{12}, \dots, \theta_{1k})$  e  $\theta_2 = (\theta_{21}, \theta_{22}, \dots, \theta_{2l})$ ,  $k, l$  conhecidos e  $\theta_1 \neq \theta_2$ . Considera-se ainda desconhecido o valor do índice  $\tau$ , onde se processa a transição, sendo pois, tido como mais um parâmetro da função de distribuição conjunta de probabilidades de  $X = (X_1, X_2, \dots, X_N)$ .

Desta forma, observada então uma sequência ordenada temporal de valores amostrais  $x_1, x_2, \dots, x_N$ , manifesta-se primeiramente interesse em fazer inferências sobre  $\tau$ . Posteriormente, e em função da problemática abordada, consegue-se também realizar inferências sobre os parâmetros  $\theta_1$  e  $\theta_2$ .

Dentro deste escopo, um grande número de casos especiais têm sido considerados. Formalizações mais complexas, acompanhadas por diferentes hipóteses de sustentação são abordadas através de diversas metodologias estatísticas, de forma a conseguir uma representação mais realista para os problemas reais específicos considerados. Shaban (1980) enumera uma vasta referência bibliográfica relativa ao assunto, classificando os trabalhos existentes através das metodologias adotadas.

Entre os casos que despertam maior interesse, situa-se o ajuste de modelos de regressão linear, compostos por diferentes segmentos funcionais, relativos a diferentes partições no domínio das variáveis independentes. Uma vasta literatura acerca do problema citado pode ser encontrada através de diversas entitulações, entre as quais cita-se "Segmented Regression", "Switching Regression", "Piecewise Regression" ou "Two Phase Regression". Para fins deste trabalho, convencionou-se denominar tal problema por "Modelo de Regressão Linear Segmentado".

Relativo a este caso especial, alguns pesquisadores fizeram uso de diferentes metodologias estatísticas em seus trabalhos, dentre os quais pode-se destacar:

Quandt (1958) retira da função de verossimilhança as estimativas dos parâmetros que compõem um modelo segmentado e descontínuo de regressão linear simples.

Sprenst (1981) analisa algumas hipóteses de interesse para o pesquisador acerca do modelo proposto por Quandt (1958), sujeito a restrição de continuidade no ponto da transição de fases. Para tanto, utiliza-se da metodologia dos mínimos quadrados restritos.

Robinson (1984) utiliza os mesmos argumentos propostos por Quandt (1958) para estimar e testar a significância dos parâmetros que compõem um modelo segmentado contínuo de regressão polinomial.

Hudson (1966) analisa, classifica e propõe alguns procedimentos de análise para o ajuste de modelos de regressão segmentado, sujeitos ou não a restrições nos parâmetros, através da metodologia dos mínimos quadrados.

Hinkley (1969, 1971) estuda o problema da estimação e testes de hipóteses acerca do ponto de transição em um modelo bi-segmentado de regressão linear simples, utilizando argumentos fundamentados na função de verossimilhança.

Gallant e Fuller (1973) consideram o problema da estimação numérica, através do método de Gauss-Newton, de modelos bi-segmentados de regressão polinomial restritos a condição de continuidade e suavidade da derivada na vizinhança do ponto de mudança de fases.

Feder (1973a, 1975b) analisa detalhadamente os procedimentos de inferência utilizados até então para testar hipóteses relativas ao ponto de transição de fases nos modelos de regressão linear segmentados. Discute também algumas dificuldades técnicas surgidas com a adoção de argumentos assintóticos baseados na função de verossimilhança.

Lerman (1980) propõe o método numérico de busca "Grid-Search", através dos argumentos de mínimos quadrados, como um procedimento alternativo àqueles propostos por Hudson (1966) para estimação do ponto de transição de fases de modelos de

regressão linear segmentados, sujeitos a restrições nos parâmetros. O método proposto por Lerman (1980) é convenientemente útil na resolução de problemas relativos ao modelo descrito por Gallant e Fuller (1973), nos quais as restrições de continuidade e suavidade da derivada na vizinhança do ponto de transição são impostas ao modelo linear segmentado.

Esterby e El-Shaarawi (1981) estabelecem um procedimento para estimação e testes de hipóteses de parâmetros de modelos polinomiais segmentados descontínuos, baseados nos conceitos de verossimilhança parcial: marginal e condicional.

Sob o ponto de vista Bayesiano, Bacon e Watts (1971), Ferreira (1975) e Salazar e Broemeling (1981) entre outros, analisaram e propuseram procedimentos de inferência para o problema. Por outro lado, Smith (1979) considerou argumentos estritamente numéricos no tratamento do problema, através de "Splines" lineares.

Uma implementação computacional de alguns procedimentos de estimação dos parâmetros de um modelo de regressão segmentado, pode ser encontrada na "procedure" NLIN do "software" estatístico SAS.

No presente trabalho, de cunho bioestatístico, procura-se ajustar um modelo de regressão linear bi-segmentado a observações temporais da resposta Ventilação Minuto Respiratória (volume de ar inspirado por minuto), obtida da experimentação em indivíduos

sadios submetidos a esforço físico progressivo. O interesse principal do trabalho consiste na estimação do ponto de transição entre os dois sub-modelos lineares considerados, objetivando a quantificação de uma característica fisiológica individual, associada ao estado físico e orgânico instantâneo dos indivíduos, denominada Limiar de Anaerobiose (LA). A partir do conhecimento deste valor, fisiologistas admitem realizar desde a avaliação das condições físicas de atletas em treinamento até a confirmação de diagnósticos em pacientes com deficiências cardiopulmonares.

No capítulo II, procura-se destacar o conceito, o mecanismo fisiológico de atuação e as formas usuais de detecção e quantificação do Limiar de Anaerobiose. Discute-se ainda, a validade destas formas de quantificação e sugere-se sua estimação através de modelos estatísticos de regressão linear bi-segmentados.

No capítulo III, trabalha-se com os modelos de regressão linear segmentados, classificando-os e apresentando procedimentos adequados para a sua análise, através do método dos mínimos quadrados ordinários.

No capítulo IV, introduz-se a noção do tratamento de modelos de regressão linear bi-segmentados, sujeitos a hipótese de resíduos autocorrelacionados, através do método dos mínimos quadrados generalizados.

Enfoca-se, visto as características do problema que se deseja estudar, apenas os procedimentos de estimação pontual do momento da transição entre os sub-modelos propostos, sem contudo, despendendo qualquer esforço com a análise de estimativas por intervalos ou testes de hipóteses acerca deste parâmetro, o que, certamente, constitui em um prosseguimento do trabalho.

No capítulo V, formula-se estatisticamente o problema biológico abordado, justificando a sua utilização. Discute-se também um estudo ilustrativo sobre o comportamento do procedimento proposto, através de alguns conjuntos de observações simuladas.

No capítulo VI, descreve-se o aparato experimental utilizado na obtenção das observações modeladas e apresenta-se os resultados obtidos através do tratamento estatístico efetuado, acompanhados por algumas discussões relevantes.

No apêndice 3, um estudo comparativo, comumente encontrado na literatura bio-médica específica, é realizado com os resultados obtidos através do modelo proposto e com os resultados obtidos através de critérios informais frequentemente utilizados por fisiologistas.



## CAPÍTULO II

### O PROBLEMA BIOLÓGICO

#### II.1. O LIMAR DE ANAEROBIOSE

A execução de trabalho muscular está associada ao incremento das atividades cardiovascular e respiratória, tendo em vista o aumento do metabolismo (Martins, 1986). Em anos recentes, muitos pesquisadores têm estudado a obtenção de um parâmetro de relacionamento entre a forma deste incremento, durante a realização de trabalho muscular, e o comprometimento do mecanismo cardiopulmonar. Dentro deste escopo, Wasserman et al. (1964) introduziram o conceito de início do Metabolismo Anaeróbico, que viria mais tarde a ser reconhecido como forma de avaliação da capacidade individual de trabalho físico.

Através de exercícios físicos de intensidade progressiva, Wasserman et al. (1964) registraram a existência de um nível de esforço, além do qual notava-se uma elevação abrupta na taxa de

ácido láctico sanguíneo. Este valor de esforço físico registrado foi interpretado como um indicador individual do início do metabolismo anaeróbico muscular e recebeu a denominação de Limiar de Anaerobiose (LA).

O conceito de Limiar de Anaerobiose (LA) tem sido usado na interpretação dos mecanismos de trocas energéticas a nível muscular, como um valor de trabalho ou como a quantidade de oxigênio consumida, a partir da qual, o processo metabólico aeróbico não consegue sustentar a produção necessária para a manutenção da concentração intramuscular de ATP (Adenosina Trifosfato), intermediário energético obrigatório na transformação da energia ingerida no alimento em energia mecânica da contração muscular.

Como consequência, o aumento do esforço muscular além do LA, provoca um incremento na glicólise anaeróbica visando suprimir o déficit intramuscular no nível de ATP, o que, por sua vez, implica na elevação da concentração muscular de ácido láctico e uma consequente acidose metabólica.

A acidose metabólica induzida pela formação de ácido láctico, provoca mudanças fisiológicas nas trocas gasosas respiratórias, estimulando o aumento excessivo da ventilação. Desta forma, o fenômeno que ocorre a nível muscular, pode ser identificado através de suas consequências a nível respiratório (Wasserman, 1986).

Inicialmente, o LA foi usado clinicamente por Wasserman et al. (1964) na avaliação da tolerância a exercícios físicos por indivíduos portadores de doenças cardiorespiratórias. Entretanto, em anos recentes, o interesse pelo LA tem se diversificado, segundo Calozzo et al. (1982), entre outros, no sentido de:

- auxiliar no diagnóstico de uma série de patologias que afetam de alguma maneira a cadeia dos processos fisiológicos do exercício;
- estudo do efeito de drogas na resistência ao exercício físico;
- prescrição de exercícios físicos;
- avaliação da evolução da capacidade física de atletas submetidos a treinamento;
- estudo de propriedades bioquímicas dos músculos;
- caracterização da resistência física de indivíduos.

Relativo a caracterização da resistência física de indivíduos, Whipp et al. (1981) descreveram o LA como um parâmetro chave na definição individual da habilidade de sustentar exercícios físicos de alta intensidade. Para os pacientes cardiopulmonares, Yeh et al. (1983) descreveram o LA como uma medida direta do esforço no qual o sistema cardiovascular falha ao tentar repor o suprimento de oxigênio necessário à manutenção do trabalho.

O uso do LA em medicina clínica, em grande parte, depende do conhecimento dos valores normais assumidos pela população

saudável. Desta forma, Naimark et al. (1964), citado por Wasserman et al. (1973), estudaram pacientes com deficiências na válvula mitral e, Wasserman et al. (1964) estudaram uma variedade de outros pacientes com doenças cardíacas, comparando e encontrando valores de LA bem abaixo dos níveis obtidos em indivíduos saudáveis.

## II.2. FORMAS DE DETECÇÃO DO LA

A existência do LA no metabolismo energético do músculo é bem aceita entre os estudiosos no assunto, entretanto, os critérios utilizados na sua detecção têm gerado controvérsias.

O processo para a obtenção do valor do LA consiste, em princípio, na análise do comportamento das respostas hematoquímicas e cardiorespiratórias a um teste ergométrico tipo rampa ou escada (Martins, 1986).

Através de métodos hematoquímicos invasivos, amostras sanguíneas são retiradas regularmente, equiespaçadamente no tempo, ao longo de experimentos com exercícios físicos progressivos, a partir das quais analisam-se as modificações no comportamento gráfico das curvas da concentração do lactato sanguíneo. De posse destas informações, segundo Yeh et al. (1983), o LA tem sido detectado visualmente, e então quantificado, a partir de diferentes critérios, entre os quais destacam-se:

- como o instante onde se observa o início de um incremento abrupto no lactato venoso;
- como o instante onde se observa o início de um crescimento não linear no lactato venoso;
- como o instante onde se observa o início de um crescimento exponencial no lactato venoso.

As variações nos critérios metodológicos de detecção visual do LA a partir de métodos invasivos, além de discordantes contam com o fator subjetivo da exata identificação visual do instante onde ocorreram tais mudanças.

Para facilitar a detecção do LA, evitando a violação do organismo do indivíduo submetido ao estudo, numerosos investigadores têm feito uso de indicadores cardiorespiratórios não invasivos, baseados na hipótese de que a acidose metabólica, induzida pela formação do ácido láctico, provocaria mudanças fisiológicas nas trocas gasosas respiratórias, ou seja, que a mudança na concentração de lactato a nível muscular e sanguíneo ocorreria quase que simultaneamente e que o ponto de elevação de lactato sanguíneo poderia ser usado para caracterizar a acidose metabólica e o limiar de um relacionamento alterado entre ventilação e consumo de oxigênio (Wasserman et al., 1973).

Segundo Yeh et al. (1983), muitos dos critérios usados para a detecção não invasiva do LA foram introduzidos por Wasserman, Whipp e Davis ao longo de suas pesquisas. O LA foi

primeiramente quantificado como sendo o nível de esforço físico instantâneo. onde se observava, visualmente, o início de um incremento substancial no Quociente Respiratório (Relação entre a eliminação de  $\text{CO}_2$  e o consumo de oxigênio). Mais tarde, defendeu-se a detecção do LA, como o nível de esforço físico instantâneo a partir do qual observava-se:

- o início de um incremento não linear na Ventilação Minuto Respiratória;
- o início de um incremento não linear na Produção de Dióxido de Carbono;

Em adição, Davis et al. (1979) propuseram também que o LA fosse quantificado como o nível de esforço físico, a partir do qual se observava o início de um incremento sistemático no Equivalente Ventilatório para  $\text{O}_2$  (isto é, Ventilação Minuto/Consumo de  $\text{O}_2$ ) sem que houvesse um incremento no Equivalente Ventilatório para  $\text{CO}_2$  (isto é, Ventilação Minuto/Produção de  $\text{CO}_2$ ). Além destes, outros tantos critérios não invasivos foram propostos e adotados por pesquisadores interessados.

### II.3. DISCUSSÕES

A grande diversidade dos critérios empregados na determinação do LA podem, segundo Martins (1986), dar uma idéia tanto da imprecisão envolvida nas medidas, devido ao carácter subjetivo, quanto da não especificidade das mesmas.

A introdução de analisadores químicos rápidos e eficientes e de computadores de processamento em tempo real têm possibilitado a detecção e a quantificação invasiva do LA com maior rapidez e confiabilidade. Por outro lado, muitos pesquisadores já utilizam o computador, também para a determinação não invasiva do parâmetro LA a partir de respostas cardiorespiratórias, contando com o auxílio de critérios matemáticos e estatísticos, o que, a princípio, implica na eliminação do carácter subjetivo embutido na determinação visual.

Dentro desta filosofia, Orr et al. (1982) descreveram um algoritmo computacional, baseado em critérios estatísticos, para detectar o instante da mudança no comportamento numérico da curva da Ventilação Minuto Respiratória versus Consumo de Oxigênio. Utilizaram-se, para tanto, das noções de modelagem estatística através de regressão linear multi-segmentada, composta por dois ou três segmentos de reta consecutivos.

Tendo em vista a importância do tema e o carácter impreciso envolvido na detecção visual, invasiva ou não, do LA, propõe-se neste trabalho a utilização das noções estatísticas de ajuste de modelos de regressão linear bi-segmentados, na tentativa de se quantificar a ocorrência de um instante de mudança no comportamento funcional de algumas medidas cardiorespiratórias, associadas à realização de um experimento envolvendo exercícios físicos progressivos, ou seja, na tentativa de se estimar o LA a partir de parâmetros de modelos estatísticos.

Deve-se notar, que a proposta de estimação estatística do LA não credibiliza o resultado obtido a responder pelo verdadeiro nível de trabalho muscular no qual ocorre o início do metabolismo anaeróbico, mas sim, representa simplesmente a formulação matemática dos critérios de detecção visual citados na literatura específica e obtidos a partir de estudos fisiológicos, desvinculados de qualquer carácter subjetivo, proveniente da experiência do pesquisador.

Para tanto, contou-se com a colaboração da Seção de Hemodinâmica e Função Pulmonar do Hospital das Clínicas FMRP-USP, onde foram realizadas sessões experimentais que culminaram na obtenção dos dados que estão analisados ao longo deste trabalho.



## CAPÍTULO III

### O MODELO DE REGRESSÃO LINEAR SEGMENTADO

#### III.1. INTRODUÇÃO

Considere um conjunto de pares de observações experimentais que aparentemente guardam entre si um relacionamento funcional, composto por uma sequência de sub-modelos lineares distintos, relativos a uma partição no domínio das variáveis independentes, associados a um componente estocástico. A tal relacionamento funcional denomina-se, estatisticamente, Modelo de Regressão Linear Segmentado.

Formalizando matematicamente esta definição, sejam  $X$  e  $Y$  duas sequências pareadas de observações experimentais que se relacionam através de um modelo de regressão linear segmentado. Portanto, considere uma certa partição no domínio da sequência  $X$  :  $\langle P_1, P_2, \dots, P_k \rangle$ , a qual induz de forma biunívoca uma partição no domínio da sequência  $Y$ . Desta forma, o relacionamento funcional

entre X e Y pode ser representado da seguinte maneira:

$$E(Y) = f_1(X) I_{p_1}(X) + f_2(X) I_{p_2}(X) + \dots + f_k(X) I_{p_k}(X)$$

onde,

$$I_{p_i}(X) = \begin{cases} 1 & \text{se } X \in P_i \\ 0 & \text{caso contrário} \end{cases}$$

$$f_i(X) = \sum_{j=0}^{G_i} \theta_{ij} x^j \quad \text{para todo } i = 1, 2, \dots, k$$

onde,  $G_i$  é o grau do polinômio de índice  $i$ .

Deste modo, tem-se que a esperança de Y,  $E(Y)$ , pode ser estimada através de uma composição de funções lineares nos parâmetros, dos valores amostrais assumidos por X, correspondentes ao elemento da partição à qual pertencem.

Através de uma notação diferente, pode-se escrever o modelo proposto como:

$$Y = \begin{cases} \sum_{j=0}^{G_1} \theta_{1j} x^j + \varepsilon_1 & ; A_0 \leq x \leq c_1 \\ \sum_{j=0}^{G_2} \theta_{2j} x^j + \varepsilon_2 & ; c_1 < x \leq c_2 \\ \dots & \\ \sum_{j=0}^{G_k} \theta_{kj} x^j + \varepsilon_k & ; c_{k-1} < x \leq A_k \end{cases} \quad (3.1)$$

onde,

$k$ ,  $A_0$ ,  $A_k$  e  $G_i$  ( $i = 1, 2, \dots, k$ ) são constantes conhecidas;

$$E(\varepsilon_i) = 0, \quad E(\varepsilon_i \varepsilon_i') = I\sigma^2; \quad i = 1, 2, \dots, k$$

Os  $x$ 's são valores amostrais distintos relativos à variável  $X$  e preservam uma certa ordenação,

$$A_0 \leq x_1 < \dots < x_{I_1} \leq c_1 < x_{I_1+1} < \dots < x_{I_k-1} \leq c_{k-1} < \dots < x_N \leq A_k$$

O vetor  $C = (c_1, c_2, \dots, c_{k-1})$  das abscissas dos pontos de transição, pode ou não ser desconhecido, sendo que, uma vez assim considerado, torna-se parâmetro do modelo, devendo ser estimado através dos dados disponíveis. Neste caso, o vetor  $I = (I_1, I_2, \dots, I_{k-1})$ , dos índices que induzem a partição na sequência ordenada, também é considerado parâmetro do modelo;

É de grande interesse o caso em que o modelo de regressão segmentado é composto por apenas dois sub-modelos lineares distintos, associados a uma partição no domínio de  $X$ , composta por apenas dois elementos,  $X : \langle P_1, P_2 \rangle$ . A este modelo denomina-se Modelo de Regressão Linear Bi-Segmentado, e pode ser representado notacionalmente como:

$$E(X) = f_1(X) I_{P_1}(X) + f_2(X) I_{P_2}(X)$$

onde,

$$I_{P_i}(X) = \begin{cases} 1 & \text{se } X \in P_i \\ 0 & \text{caso contrário} \end{cases}$$

$$f_i(X) = \sum_{j=0}^{G_i} \theta_{ij} x^j$$

onde,  $G_i$  é o grau do polinômio de índice  $i$  ( $i = 1, 2$ ).

ou ainda,

$$Y = \begin{cases} \sum_{j=0}^{G_1} \theta_{1j} x^j + \varepsilon_1 & ; A_0 \leq x \leq c_1 \\ \sum_{j=0}^{G_2} \theta_{2j} x^j + \varepsilon_2 & ; c_1 < x \leq A_2 \end{cases} \quad (3.2)$$

Para o qual supõe-se todas as considerações feitas ao modelo segmentado (3.1), mantida a devida dimensionalidade.

Seria relativamente simples se o modelo definido em (3.1) ou (3.2) levasse em consideração o prévio conhecimento dos valores das abscissas dos pontos de transição (no caso, o vetor  $C$ ). Porém, frequentemente, depara-se com o caso em que a falta de conhecimento prévio deste vetor, obriga o pesquisador a estimá-lo a partir dos dados disponíveis, o que implica em um substancial aumento no grau de dificuldade analítica do problema.

Em associação às dificuldades proporcionadas pelo desconhecimento dos valores do vetor  $C$ , muitas vezes depara-se com a necessidade de se impor determinadas restrições aos parâmetros  $\theta$  do modelo, visando uma melhor adequação do problema estatístico formalizado ao fenômeno real em análise, implicando assim, novamente, no crescimento da complexidade do modelo proposto.

Tendo em vista o modelo abordado (3.1), deseja-se portanto encontrar estimativas para os parâmetros  $\theta_1, \theta_2, \dots, \theta_k, C$  e  $I$ . Para tanto, comumente, faz-se uso do critério dos Mínimos Quadrados, ou seja, estimam-se os parâmetros  $\theta_1, \theta_2, \dots, \theta_k, C$  e  $I$ , tais que minimizem a função Soma das Somas de Quadrados de Resíduos (SSQR), dada por:

$$SSQR = \sum_{i=1}^k \sum_{c_{i-1} < x_j \leq c_i} [y_j - f_i(x_j, \theta_i)]^2$$

onde  $co = A_0$  e  $ck = A_k$

sujeita, ou não, a determinadas restrições nos parâmetros.

No caso de modelos de regressão linear bi-segmentados, os estimadores de mínimos quadrados para  $\theta_1$ ,  $\theta_2$ ,  $c$  e  $I$  ( $\hat{\theta}_1$ ,  $\hat{\theta}_2$ ,  $\hat{c}$  e  $\hat{I}$ ) são valores que minimizam a função Soma das Somas de Quadrados de resíduos, dada por:

$$SSQR = \sum_{j=1}^I [y_j - f_1(x_j, \theta_1)]^2 + \sum_{j=I+1}^N [y_j - f_2(x_j, \theta_2)]^2$$

sujeita, ou não, a determinadas restrições nos parâmetros.

A adoção do critério exposto na estimação analítica dos parâmetros do modelo tratado, torna-se um trabalho não trivial, visto que a função SSQR possui, potencialmente, diversos mínimos locais e é não diferenciável em um certo número de pontos (especificamente se algum  $x_j$  coincidir com alguma das abscissas dos pontos de transição ci)

Trata-se portanto de um trabalho estatístico de razoável complexidade, porém, de grande aplicabilidade em problemas reais, o que contribui para o prosseguimento das pesquisas a ele dedicadas.

Convém salientar que a partir deste momento, o trabalho será restrito ao tratamento de modelos de regressão linear bi-segmentados (3.2), tendo em vista a resolução do problema biológico que, ao final, merecerá consideração.

### III.2. CLASSIFICAÇÃO DOS MODELOS DE REGRESSÃO LINEAR BI-SEGMENTADOS

Hudson (1966) sugere uma classificação para os modelos de regressão linear bi-segmentados através das restrições impostas aos parâmetros, adotando procedimentos óptimos de resolução e análise para cada tipo de modelo sugerido.

Nesta forma, distinguem-se os modelos de regressão linear bi-segmentados contínuos, através da imposição da seguinte restrição aos parâmetros do modelo:

$$A: f_1(\hat{c}, \hat{\theta}_1) = f_2(\hat{c}, \hat{\theta}_2) \quad (3.3)$$

No caso de modelos contínuos, o ponto  $c$  é denominado ponto de junção dos sub-modelos. No caso de modelos descontínuos, a referência ao ponto  $c$  é feita como ponto de mudança ou troca de fases.

Antes de retomar a sequência da explanação, comenta-se que o ajuste de modelos de regressão linear bi-segmentados descontínuos é um problema de fácil resolução mas, contudo, não será abordado neste trabalho (ver Hawkins, 1976).

Cogita-se ainda, agora relativo aos modelos contínuos, sobre a possibilidade do ponto de junção  $c$  coincidir ou não com um dos valores amostrais observados  $x$ 's e, sobre a forma como ocorre essa junção, suave ou bruscamente. Estatisticamente, tais conjecturas podem ser incorporadas ao modelo de regressão linear bi-segmentado, através das seguintes restrições impostas aos parâmetros:

$$B: \hat{c} = x_I \quad \text{ou} \quad x_I < \hat{c} < x_{I+1} \quad (3.4)$$

$$C: f'_1(\hat{c}, \hat{\theta}_1) = f'_2(\hat{c}, \hat{\theta}_2) \quad (3.5)$$

onde,  $\hat{I} = (G1+1, G1+2, \dots, N-(G2+1))$ ;

$G1+1$  é o número de parâmetros a serem estimados no sub-modelo  $f_1(\dots)$ ;

$G2+1$  é o número de parâmetros a serem estimados no sub-modelo  $f_2(\dots)$ ;

$$f'_i(\hat{c}, \hat{\theta}) = \partial f_i(X, \hat{\theta}) / \partial X \Big|_{x=\hat{c}}$$

Assim, apesar de tratarem-se todos, de modelos de regressão linear bi-segmentados, contínuos e com restrições nos parâmetros, Hudson (1966) classifica-os em três tipos, visando otimizar os procedimentos de análise. Segue-se a tabela,

TABELA III.1. Classificação dos modelos de regressão linear bi-segmentados segundo Hudson.

	$\hat{c} \neq x \hat{I}$	$\hat{c} = x \hat{I}$
$f'_1(\hat{c}, \hat{\theta}_1) \neq f'_2(\hat{c}, \hat{\theta}_2)$ (junção brusca)	TIPO I	TIPO II
$f'_1(\hat{c}, \hat{\theta}_1) = f'_2(\hat{c}, \hat{\theta}_2)$ (junção suave)	TIPO III	TIPO II



Detalhando um pouco mais as informações contidas na Tabela (III.1.), segue-se:

1. Junção do TIPO I. Uma junção é dita do TIPO I quando:

(i)  $\hat{c}$  está estritamente restrito ao intervalo entre dois valores sucessivos  $x$ 's, isto é,

$$x_{\hat{I}} < \hat{c} < x_{\hat{I}+1} ; \quad \hat{I} = G1+1, G1+2, \dots, N-(G2+1)$$

(ii) as declividades dos modelos adjacentes não são iguais no ponto de junção, isto é,

$$f'_1(\hat{c}, \hat{\theta}_1) \neq f'_2(\hat{c}, \hat{\theta}_2)$$

2. Junção do TIPO II. Uma junção é dita do TIPO II quando:

(i)  $\hat{c}$  é igual a uma das observações amostrais  $x$ 's, isto é,

$$\hat{c} = x_{\hat{I}} \quad \hat{I} = G1+1, G1+2, \dots, N-(G2+1)$$

3. Junção do TIPO III. Uma junção é dita do tipo III quando:

(i)  $\hat{c}$  está estritamente restrito ao intervalo entre dois valores sucessivos  $x$ 's, isto é,

$$x_{\hat{I}} < \hat{c} < x_{\hat{I}+1} ; \quad \hat{I} = G1+1, G1+2, \dots, N-(G2+1)$$

(ii) as declividades dos modelos adjacentes são iguais no ponto de junção, isto é,

$$f'_1(\hat{c}, \hat{\theta}_1) = f'_2(\hat{c}, \hat{\theta}_2)$$

Neste caso, deve-se observar que  $f_1(\dots)$  e  $f_2(\dots)$  não podem ter, simultaneamente, as formas funcionais de uma constante ou de uma reta, ressalva-se o caso em que  $f_1(\dots) = f_2(\dots)$ .

### III.3. PROCEDIMENTOS DE ESTIMAÇÃO

Classificados os modelos que porventura possam vir a ser considerados sob o enfoque de regressão linear bi-segmentados, apresentam-se, nesta etapa, alguns procedimentos estatísticos relativos a estimação dos parâmetros  $\theta_1$ ,  $\theta_2$ ,  $c$  e  $I$ , através do critério dos mínimos quadrados, de forma a obter um mínimo global para a função Soma das Somas de Quadrados de Resíduos (SSQR), dada por:

$$SSQR = \sum_{j=1}^I [y_j - f_1(x_j, \theta_1)]^2 + \sum_{j=I+1}^N [y_j - f_2(x_j, \theta_2)]^2 \quad (3.6)$$

levando em consideração algumas das relações de restrição às estimativas dos parâmetros, apresentadas em (3.3), (3.4) e (3.5).

A princípio, suponha conhecidas as formas funcionais polinomiais (G1 e G2) das funções  $f_1(\dots)$  e  $f_2(\dots)$ . Sendo assim, distinguem-se dois casos relativos a solução do problema: o primeiro caracterizado pelo conhecimento prévio do valor do ponto de junção dos modelos adjacentes,  $c$ , e o segundo, caracterizado pelo desconhecimento deste valor e pelo desejo de estimá-lo através das observações amostrais disponíveis. Segue-se uma descrição da forma de atuação nos dois casos citados.

### CASO 1. Ponto de junção conhecido.

Suponha conhecido o ponto  $c$  de junção dos sub-modelos adjacentes e, em consequência, o índice  $I$  que determina a partição no conjunto ordenado das variáveis independentes  $X$ . O procedimento de estimação dos parâmetros  $\theta_1$  e  $\theta_2$  se restringe basicamente ao ajuste, em separado, via mínimos quadrados restritos (ver apêndice 2), dos sub-modelos de regressão adjacentes. Para tanto, basta ajustar o primeiro sub-modelo  $f_1(X, \theta_1)$  para os primeiros  $I$  pares de observações  $(x_1, y_1), (x_2, y_2), \dots, (x_I, y_I)$  e, o segundo sub-modelo  $f_2(X, \theta_2)$  para os restantes  $(N-I)$  pares de observações  $(x_{I+1}, y_{I+1}), (x_{I+2}, y_{I+2}), \dots, (x_N, y_N)$ , condicionados às restrições impostas aos parâmetros.

Por exemplo, suponha o ajuste do modelo considerado a seguir:

$$y_j = \begin{cases} \theta_{11} + \theta_{12} x_j + \varepsilon_j & ; j = 1, 2, \dots, I \\ \theta_{21} + \theta_{22} x_j + \varepsilon_j & ; j = I+1, I+2, \dots, N \end{cases} \quad (3.7)$$

sujeito a restrição de continuidade no ponto  $X = c$ , isto é,

$$\theta_{11} + \theta_{12} c = \theta_{21} + \theta_{22} c$$

Sejam as observações amostrais, representadas sob a forma matricial da seguinte forma:

$$Y_{N \times 1} = (y_1 \ y_2 \ \dots \ y_I \ y_{I+1} \ \dots \ y_N)'$$

$$\theta_{4 \times 1} = (\theta_{11} \ \theta_{12} \ \theta_{21} \ \theta_{22})'$$

$$\varepsilon_{N \times 1} = (\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_I \ \varepsilon_{I+1} \ \dots \ \varepsilon_N)'$$

$$X_{N \times 4} = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 1 & x_I & 0 & 0 \\ 0 & 0 & 1 & x_{I+1} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 1 & x_{N-1} \\ 0 & 0 & 1 & x_N \end{pmatrix}$$

$$R_{1 \times 4} = (1 \ c \ -1 \ -c)$$

Desta maneira, obtém-se o estimador  $\hat{\theta}^+$  de  $\theta$  (apêndice 2) através da expressão:

$$\hat{\theta}^+ = \hat{\theta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\hat{\theta}$$

onde,  $\hat{\theta} = (X'X)^{-1}X'Y$

## CASO 2. Ponto De Junção Desconhecido.

Suponha agora desconhecido o valor do ponto  $c$  de junção dos modelos adjacentes. Neste caso, deve-se conseguir estimá-lo através das observações amostrais disponíveis.

Como se supõe  $\hat{c}$  e  $[x_1^0, x_{I+1}^0]$ , o modelo torna-se não linear nos parâmetros e a sua estimação analítica por mínimos quadrados torna-se bastante complexa, visto que, tanto  $c$  quanto  $I$  são considerados parâmetros adicionais ao modelo tratado.

Hudson (1966) comenta que a resolução analítica deste modelo pode apresentar grandes dificuldades algébricas e, portanto, propõe que a solução do problema seja pesquisada através de técnicas iterativas.

De acordo com a sugestão de Hudson (1966), Lerman (1980) propõe um procedimento "alternativo" embasado em conceitos de uso frequente na estimação de parâmetros de modelos de regressão não linear.

Segue-se uma descrição sucinta do procedimento proposto por Hudson (1966) e do procedimento proposto por Lerman (1980), seguida de alguns comentários acerca das vantagens ou desvantagens quando do uso de um ou outro critério.

### III.3.1 O PROCEDIMENTO DE ESTIMAÇÃO DE HUDSON

O Procedimento "padrão" (denominação de Lerman, 1980), descrito por Hudson (1966) e mais tarde refinado por Hinkley (1969, 1971), leva em consideração a classificação do problema abordado segundo a sua proposta, apresentada na Tabela III.1.

Desta forma, a cada um dos três tipos de modelos descritos, associa-se um procedimento específico para a estimação dos parâmetros  $c$  e  $I$  do modelo de regressão linear bi-segmentado, visando a otimização do trabalho computacional requerido.

Tais procedimentos partem do princípio de que o pesquisador interessado na aplicação da metodologia possui o conhecimento prévio acerca da classificação (segundo a Tabela III.1.) de seu problema específico. Desta forma, ele poderá recorrer diretamente a um dos três tipos de procedimentos propostos e conseguir, de forma otimizada, os resultados desejados.

Entretanto, se não for possível classificar previamente o problema em questão, Hudson (1966) cita ainda um procedimento geral para a estimação desses parâmetros.

Desta forma, exetutando-se o procedimento proposto relativo a modelos do TIPO III, que merecem atenção especial, o procedimento geral sugerido por Hudson está baseado em dois estágios:

1. No primeiro estágio, ajusta-se o modelo (3.2), através das expressões de mínimos quadrados ordinários, para toda partição das observações, isto é, para todo  $I=G1+1, G1+2, \dots, N-(G2+1)$ , sem que qualquer restrição seja imposta aos parâmetros.

Para cada partição tomada, as estimativas dos pontos de transição  $\hat{c}$ , são obtidas através de uma função linear das estimativas  $\hat{\theta}_1$  e  $\hat{\theta}_2$ , resultante da formulação algébrica da restrição de continuidade, ou seja,  $f_1(\hat{c}, \hat{\theta}_1) = f_2(\hat{c}, \hat{\theta}_2)$ .

Por exemplo, caso o modelo considerado fosse aquele descrito em (3.7), o valor da estimativa  $\hat{c}$ , seria obtido das estimativas  $\hat{\theta}_{11}$ ,  $\hat{\theta}_{12}$ ,  $\hat{\theta}_{21}$  e  $\hat{\theta}_{22}$ , através da formulação algébrica da restrição de continuidade,

$$\hat{\theta}_{11} + \hat{\theta}_{12} \hat{c} = \hat{\theta}_{21} + \hat{\theta}_{22} \hat{c} \quad \text{que resulta em,}$$

$$\hat{c} = (\hat{\theta}_{21} - \hat{\theta}_{11}) / (\hat{\theta}_{12} - \hat{\theta}_{22})$$

Se esta estimativa de ponto de junção obtida,  $\hat{c}$ , dividir as observações na mesma partição fixada antes do ajuste (isto é se  $x_1^* < \hat{c} < x_{I+1}^*$ ) então o ajuste é dito admissível, e sua SSQR é anotada.

Hudson prova que, quando o ajuste é dito admissível, a solução do modelo segmentado irrestrito é igual à solução do modelo segmentado restrito.

2. No segundo estágio, retomam-se aquelas partições viáveis cujos ajustes não foram considerados admissíveis no primeiro estágio, submetendo o modelo proposto ao método de estimação discutido no apêndice 2 (Solução do Modelo de Regressão Linear com Restrições nos Parâmetros), levando em consideração, desta vez, a restrição de que o ponto de junção deve coincidir com uma das abscissas limitantes do intervalo considerado ( $\hat{c} = x_I^0$  ou  $\hat{c} = x_{I+1}^0$ ), além das demais restrições que porventura venham a ser consideradas. Desta vez, procede-se a anotação do valor da estimativa  $\hat{SSQR}^*$ .

A solução de mínimos quadrados para o problema de estimação dos parâmetros  $\theta_1$ ,  $\theta_2$ ,  $c$  e  $I$  do modelo de regressão linear bi-segmentado (segundo o procedimento de Hudson) é então tomada relativamente ao mínimo valor numérico estimado de  $\hat{SSQR}$  ou  $\hat{SSQR}^*$ , obtido nos dois estágios descritos.

### III.3.2. O PROCEDIMENTO DE ESTIMAÇÃO DE LERMAN

Um procedimento alternativo àquele proposto por Hudson (1966) foi apresentado por Lerman (1980) e está baseado em uma adaptação da técnica do "Grid-Search", utilizada na resolução de modelos de regressão não lineares.

Este, consiste do mapeamento da função  $\hat{SSQR}$  (3.6) sobre toda uma série de abscissas  $\hat{c}$ 's pré-fixadas, prováveis



candidatas à estimativa do ponto de junção  $c$ , obtidas através da divisão dos intervalos  $[x_I^{\wedge}, x_{I+1}^{\wedge}]$  em sub-intervalos menores. Desta forma, embasado na proposta de Lerman, de carácter computacionalmente intensivo, adota-se o seguinte procedimento na solução do problema em questão:

1. Cada intervalo  $[x_I^{\wedge}, x_{I+1}^{\wedge}]$  onde provavelmente possa ocorrer o ponto de junção, deve ser dividido em  $k$  partes, onde  $k$  dependerá da amplitude do intervalo e da precisão com que se deseja estimar  $c$ . Portanto toma-se,

$$d = (x_{I+1}^{\wedge} - x_I^{\wedge}) / k$$

2. Ao limite inferior do intervalo,  $x_I^{\wedge}$ , adiciona-se o valor do incremento  $d$  calculado no passo 1, de forma a obter  $k-1$  valores intermediários.

$$x_I^{\wedge} < x_{I+d}^{\wedge} < x_{I+2d}^{\wedge} < \dots < x_{I+(k-1)d}^{\wedge} < x_{I+1}^{\wedge}$$

3. Para cada valor obtido no passo anterior (cada nova abcissa fixada, correspondente a um provável ponto de junção  $c$ ), determina-se  $\hat{\theta}_1^+$ ,  $\hat{\theta}_2^+$  e também o valor da estimativa da função SSQR, tal como no CASO 1 (sessão III.3.), onde conhecia-se antecipadamente o valor de  $c$  e  $I$ .

4. Finalmente, adota-se como solução do problema de estimação dos parâmetros  $\theta_1$ ,  $\theta_2$ ,  $c$  e  $I$ , as estimativas correspondentes à menor  $\hat{SSQR}^+$ , sobre todos os possíveis  $\hat{I}$ .

Comenta-se que o procedimento proposto por Hudson é devidamente apropriado e pode ser totalmente implementado computacionalmente, visto tratar-se de um procedimento algorítmico. Entretanto, deixa bastante a desejar ou simplesmente é dito inapropriado em certos casos mais gerais, como por exemplo quando o modelo segmentado em estudo compõe-se de sub-modelo não lineares nos parâmetros, ou então quando, mesmo composto por sub-modelos lineares, existirem razões contundentes para se definir prévios intervalos acerca da estimação do ponto de junção.

Relativo à consideração de sub-modelos não lineares, pode-se dizer que o procedimento de Hudson é inapropriado, o que, no procedimento de Lerman pode ser facilmente incorporado simplesmente mediante a modificação da metodologia de estimação citada (seção III.3.2 - passo 3) por outra devidamente conveniente, como por exemplo, a técnica de Marquardt (ver Draper & Smith, 1966).

Facilmente, também, incorporam-se ao procedimento alternativo de Lerman certas informações disponíveis acerca da ocorrência do valor desconhecido do ponto de junção  $c$ , ao se considerar os possíveis  $\hat{I}$ 's fixados antecipadamente. Neste caso, seria conveniente abordar o assunto através da metodologia Bayesiana, que pode ser fartamente encontrada na literatura específica (ver Ferreira, 1975, entre outros) mas foge do escopo deste trabalho.

## CAPÍTULO IV

### AUTOCORRELAÇÃO SERIAL E MÍNIMOS QUADRADOS GENERALIZADOS

#### IV.1. INTRODUÇÃO

Uma das hipóteses postuladas na apresentação do modelo de regressão linear bi-segmentado (3.2), é a independência serial do termo de resíduo aleatório, implícita em:

$$E(\varepsilon\varepsilon') = \sigma^2 \quad \text{ou}$$

$$E(\varepsilon_i \varepsilon_{i+s}) = 0 \quad \text{para todo } i \ (i=1,2,\dots,N) \text{ e todo } s \neq 0.$$

Desta maneira, supunha-se que os sucessivos valores de resíduos comportavam-se independentemente de seus valores prévios.

Existem entretanto, situações nas quais a hipótese de independência serial do termo de resíduo pode não ser plausível.

Por exemplo, geralmente quando as observações amostrais tratam-se de séries cronológicas de dados, onde cada valor corresponde a um período de tempo (ano, mês, ...), ou também quando ocorre uma especificação incorreta da forma funcional da relação entre as variáveis, ou mesmo talvez quando ocorrem erros de medida nas variáveis observadas.

Caso as estimativas  $\hat{\theta}$  de mínimos quadrados ordinários forem obtidas diretamente de conjuntos de observações sobre os quais não se pode postular a afirmação de independência serial dos resíduos, sujeitam-se às seguintes consequências:

- obtenção de estimativas não tendenciosas para os parâmetros  $\theta$  do modelo;
- obtenção de estimativas tendenciosas da variância dos resíduos e, portanto, invalidação dos testes  $t$  de Student e  $F$  de Fisher, realizados sobre a significância dos parâmetros  $\theta$  do modelo;
- obtenção de previsões ineficientes, com variâncias amostrais desnecessariamente grandes.

A seguir aborda-se tal problema sob o contexto do método dos Mínimos Quadrados Generalizados, a partir do qual esboçam-se algumas das provas das situações comentadas acima.

## IV.2. O MÉTODO DOS MÍNIMOS QUADRADOS GENERALIZADOS

Para uma melhor abordagem deste problema, seja o modelo de regressão a seguir,

$$Y = X\beta + v \quad (4.1)$$

onde,

$$Y = (y_1, y_2, \dots, y_N)'$$

$$\beta = (\beta_1, \beta_2, \dots, \beta_k)'$$

$$v = (v_1, v_2, \dots, v_N)'$$

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ & & \dots & \\ x_{N1} & x_{N2} & \dots & x_{Nk} \end{pmatrix}$$

Suponha agora que

$$E(v) = 0 \quad \text{e} \quad E(vv') = V\sigma^2,$$

onde  $V$ , a matriz de variância-covariância do termo de resíduo, é uma matriz  $N \times N$ , simétrica, positiva definida e conhecida; então  $V^{-1}$ , a matriz inversa de  $V$ , é uma matriz  $N \times N$ , simétrica, positiva definida e existe uma matriz não-singular  $P$ , tal que:

$$P'P = V^{-1} \quad (4.2)$$

Pré-multiplicando (4.1) por P, tem-se,

$$PY = PX\beta + Pu \quad (4.3)$$

Para o modelo de regressão (4.3) de PY contra PX, define-se  $e = Pu$ , e então calcula-se,

$$E(e) = E(Pu) = PE(u) = 0 \quad e$$

$$\begin{aligned} E(ee') &= E(Puu'P') = PVP'\sigma^2 \\ &= P(P'P)^{-1}P'\sigma^2 \\ &= PP^{-1}P^{-1}P'\sigma^2 \\ &= I\sigma^2 \end{aligned}$$

Desta forma, pode-se portanto, aplicar ao modelo de regressão (4.3) as expressões conhecidas de mínimos quadrados ordinários, obtendo,

$$\hat{\beta}^* = (X'P'PX)^{-1}X'P'PY = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad (4.4)$$

que é o melhor estimador linear, não tendencioso de variância mínima de  $\beta$ , ou seja,  $\hat{\beta}^*$  é BLUE (Teorema de Aitken).

A demonstração desse teorema pode ser obtida através da seguinte sequência de passos:

$$\begin{aligned}
1. \quad E(\hat{\beta}^*) &= E[(X'V^{-1}X)^{-1}X'V^{-1}Y] \\
&= (X'V^{-1}X)^{-1}X'P' \cdot E(PY) \\
&= (X'V^{-1}X)^{-1}X'P'PX\beta \\
&= (X'V^{-1}X)^{-1}X'V^{-1}X\beta \\
&= \beta
\end{aligned}$$

$$\begin{aligned}
2. \quad \hat{\beta}^* &= (X'V^{-1}X)^{-1}X'V^{-1}Y \\
&= (X'V^{-1}X)^{-1}X'P'PY \\
&= (X'V^{-1}X)^{-1}X'P'(PX\beta + Pv) \\
&= (X'V^{-1}X)^{-1}X'V^{-1}X\beta + (X'V^{-1}X)^{-1}X'V^{-1}v \\
&= \beta + (X'V^{-1}X)^{-1}X'V^{-1}v
\end{aligned}$$

$$\text{e portanto, } \hat{\beta}^* - \beta = (X'V^{-1}X)^{-1}X'V^{-1}v$$

de posse dessa identidade algébrica, denominada "erro amostral", obtém-se a variância do estimador:

$$\begin{aligned}
V(\hat{\beta}^*) &= E[(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)'] \\
&= E[(X'V^{-1}X)^{-1}X'V^{-1}vv'V^{-1}X(X'V^{-1}X)^{-1}] \\
&= (X'V^{-1}X)^{-1}X'V^{-1} \cdot E(vv') \cdot V^{-1}X(X'V^{-1}X)^{-1} \\
&= (X'V^{-1}X)^{-1}X'V^{-1}VV^{-1}X(X'V^{-1}X)^{-1}\sigma^2 \\
&= (X'V^{-1}X)^{-1}\sigma^2
\end{aligned} \tag{4.5}$$

$$3. \text{ Seja agora, } [(X'V^{-1}X)^{-1}X'V^{-1} + A]Y$$

um estimador linear qualquer de  $\beta$ , onde  $A$  é uma matriz  $K \times N$ .

$$[(X'V^{-1}X)^{-1}X'V^{-1} + A]Y =$$

$$[(X'V^{-1}X)^{-1}X'V^{-1} + A](X\beta + v) =$$

$$= \beta + (X'V^{-1}X)^{-1}X'V^{-1}u + AX\beta + Au$$

para que esse novo estimador seja não tendencioso, necessita-se que:

$$AX = 0 \quad (4.6)$$

portanto, o erro amostral desse estimador é dado por:

$$[(X'V^{-1}X)^{-1}X'V^{-1} + A]u$$

e a matriz de covariâncias,

$$\begin{aligned} & [(X'V^{-1}X)^{-1}X'V^{-1} + A].E(uu').[(X'V^{-1}X)^{-1}X'V^{-1} + A]' = \\ & \sigma^2[(X'V^{-1}X)^{-1}X'V^{-1} + A]V[(X'V^{-1}X)^{-1}X'V^{-1} + A]' = \\ & \sigma^2(X'V^{-1}X)^{-1} + \sigma^2AVA' + \sigma^2(X'V^{-1}X)^{-1}X'A' + \sigma^2AX(X'V^{-1}X)^{-1} \end{aligned}$$

os dois últimos termos são nulos, visto a observação feita em (4.6), então, essa matriz de covariâncias excede  $V(\hat{\beta}^*)$  (4.5), por uma matriz positiva semi-definida  $\sigma^2AVA'$ .

Vejamos a seguir, o que aconteceria se, erroneamente, aplicássemos mínimos quadrados ordinários ao modelo (4.1). Admitamos portanto, que para tal modelo fosse calculado o estimador OLS:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + u) \\ &= \beta + (X'X)^{-1}X'u \end{aligned}$$



Uma vez que  $EC(u) = 0$ , então,  $E(\hat{\beta}) = \beta$ , isto é, o estimador OLS é não tendencioso para os parâmetros  $\beta$  do modelo (4.1). Segue-se que  $(\hat{\beta} - \beta) = (X'X)^{-1}X'u$ , e portanto,

$$\begin{aligned} E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] &= [(X'X)^{-1}X' \cdot E(uu') \cdot X(X'X)^{-1}] \\ &= \sigma^2 [(X'X)^{-1}X' V X(X'X)^{-1}] \end{aligned}$$

que, no caso de  $V = I$  seria igual a  $\sigma^2(X'X)^{-1}$ .

Assim, como  $V \neq I$ , o cálculo das estimativas de  $\beta$  através de mínimos quadrados ordinários,  $\hat{\beta}$ , resultaria em estimativas não eficientes.

Dando prosseguimento ao estudo do método dos mínimos quadrados generalizados, o cálculo da estimativa da soma de quadrados de resíduos é dada por:

$$\begin{aligned} \hat{SQR}^* &= \hat{\varepsilon}^{*'} \hat{\varepsilon}^* = (PY - PX\hat{\beta}^*)(PY - PX\hat{\beta}^*) \\ &= (Y - X\hat{\beta}^*)(Y - X\hat{\beta}^*)' V^{-1} (Y - X\hat{\beta}^*) \\ &= Y' V^{-1} Y - 2Y' V^{-1} X \hat{\beta}^* + \hat{\beta}^{*'} X' V^{-1} X \hat{\beta}^* \end{aligned}$$

lembrando, em (4.4), que  $\hat{\beta}^{*'} = Y' V^{-1} X (X' V^{-1} X)^{-1}$

portanto,

$$\begin{aligned} \hat{SQR}^* &= Y' V^{-1} Y - 2Y' V^{-1} X \hat{\beta}^* + Y' V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1} X \hat{\beta}^* \\ &= Y' V^{-1} Y - Y' V^{-1} X \hat{\beta}^* \end{aligned} \tag{4.7}$$

Finalizando, como informação adicional, comenta-se que a estatística,

$$\hat{\sigma}^{2*} = (Y - X\hat{\beta}^*)' V^{-1} (Y - X\hat{\beta}^*) / (N-p)$$

onde  $p = k + 1$ ,

é um estimador não tendencioso de  $\sigma^2$ .

#### IV.3. MÍNIMOS QUADRADOS GENERALIZADOS E RESÍDUOS AUTOCORRELACIONADOS

Seja o modelo de regressão,

$$Y = X\beta + v \quad \text{ou} \quad y_t = \sum_{i=0}^k \beta_i x_{it} + v_t \quad (4.8)$$

$$\text{com } v_t = \rho v_{t-1} + \varepsilon_t \quad t = 1, 2, \dots, N$$

onde  $|\rho| < 1$  e  $\varepsilon_t$  satisfazem às seguintes hipóteses:

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t^2) = \sigma^2$$

$$E(\varepsilon_t \varepsilon_{t-h}) = 0 \quad \text{para todo } h \neq 0 \quad ; \quad t = 1, 2, \dots, N$$

Seja  $x_{0t} = 1$  para todo  $t = 1, 2, \dots, N$ . Assim, o posto da matriz  $X$ ,  $\text{Rank}(X)$ , é igual a  $p = k + 1$ . Tem-se então,

$$\begin{aligned}
v_t &= \rho v_{t-1} + \varepsilon_t \\
&= \rho (\rho v_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \dots \\
&= \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \dots
\end{aligned}$$

isto é,  $v_t = \sum_{\tau=0}^{\infty} \rho^{\tau} \varepsilon_{t-\tau}$

portanto,  $E(v_t) = 0$

$$\begin{aligned}
V(v_t) &= E(v_t^2) = (1 + \rho^2 + \rho^4 + \dots) \sigma^2 \\
&= \sigma^2 / (1 - \rho^2)
\end{aligned}$$

$$\begin{aligned}
E(v_t v_{t-h}) &= E[(\varepsilon_t + \rho \varepsilon_{t-1} + \dots)(\varepsilon_{t-h} + \rho \varepsilon_{t-h-1} + \dots)] \\
&= \rho^h \sigma^2 + \rho^{h+2} \sigma^2 + \rho^{h+4} \sigma^2 + \dots \\
&= \sigma^2 \rho^h (1 + \rho^2 + \rho^4 + \dots) \\
&= \sigma^2 \rho^h / (1 - \rho^2)
\end{aligned}$$

portanto,  $E(v) = 0$  e  $E(vv') = V\sigma^2$ , onde:

$$V = (1 - \rho^2)^{-1} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-3} \\ & & & \dots & \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{bmatrix}$$

$$V^{-1} = P'P = \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}$$

e.

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}$$

e desta forma, conhecido o valor de  $\rho$ , pode-se aplicar as expressões de mínimos quadrados generalizados vistas até então, para a estimação dos parâmetros  $\beta$  do modelo (4.8).

#### IV.3.1. MÉTODOS DE ESTIMAÇÃO DO PARÂMETRO AUTORREGRESSIVO

Segue-se uma descrição de alguns procedimentos de estimação do parâmetro autorregressivo  $\rho$ , o que consiste em uma etapa preliminar no ajuste do modelo (4.8), através de mínimos quadrados generalizados.

## 1. Procedimentos "Two-Step":

- O primeiro passo envolve a estimação de  $\rho$  e,
- o segundo passo prepara a matriz  $V$  e estima  $\beta$  através das expressões de mínimos quadrados generalizados (4.4).

O procedimento pode ser repetido iterativamente, até que sucessivos valores de  $\hat{\rho}$  sejam tão próximos quanto uma certa medida de erro adotada (isto é, até que sejam aproximadamente iguais).

### 1.1. Procedimento de Cochrane-Orcutt.

Estima-se  $\beta$  por mínimos quadrados ordinários; tomam-se os resíduos estimados,  $\hat{u}_t$ 's, e ajusta-se uma regressão linear simples de  $\hat{u}_t$  contra  $\hat{u}_{t-1}$ ; estima-se  $\rho$  através do valor ajustado do parâmetro  $\hat{u}_{t-1}$ , isto é:

$$\hat{\rho} = \left( \sum_{t=2}^N \hat{u}_t \hat{u}_{t-1} \right) / \sum_{t=1}^N \hat{u}_t^2$$

### 1.2. Procedimento de Durbin.

Estima-se uma equação de regressão de  $y_t$  contra  $y_{t-1}$ ,  $x_t$  e  $x_{t-1}$ ; toma-se a estimativa do coeficiente de  $y_{t-1}$  como a estimativa de  $\rho$ .

Comenta-se que o procedimento de Cochrane-Orcutt envolve um esforço computacional menor do que o procedimento de Durbin,

principalmente quando é grande o número de variáveis independentes.

## 2. Procedimento "Search" de Hildreth e Lu.

Calcula-se  $\hat{\beta}^*$  por mínimos quadrados generalizados para diferentes valores de  $\hat{\rho}$  em intervalos de comprimento 0.1 no intervalo  $[-1,1]$ ; calcula-se a  $SQR^*$  (4.7) em cada caso; escolhem-se aquelas estimativas de  $\hat{\rho}$  e de  $\hat{\beta}^*$  correspondentes à menor  $SQR^*$ . Repete-se o processo para intervalos menores em torno da estimativa  $\hat{\rho}$  escolhida.

## IV.4. MÍNIMOS QUADRADOS GENERALIZADOS E RESTRIÇÕES LINEARES NOS PARÂMETROS

Dando prosseguimento ao estudo da metodologia comumente utilizada no ajuste de modelos de regressão linear na presença de resíduos autocorrelacionados, apresenta-se nesta seção uma adaptação, via mínimos quadrados generalizados, dos resultados obtidos no apêndice 2, quando da estimação dos parâmetros de um modelo de regressão sujeitos a restrições lineares.

Desta forma, seja o modelo,

$$Y = X\beta + v \quad \text{sujeito às restrições,} \quad (4.9)$$

$$R\beta = 0$$

onde,  $EC(v) = 0$  e  $EC(vv') = V\sigma^2$

Devemos então estimar os parâmetros  $\beta$  do modelo (4.9), através de  $\hat{\beta}^{\#}$ , de forma a minimizar a função  $SQR^{\#} = (Y - X\beta)'V^{-1}(Y - X\beta)$ , sujeitos às restrições  $R\beta = 0$ .

Seja a função  $S = (Y - X\beta)'V^{-1}(Y - X\beta) + 2\lambda'R\beta$  onde  $\lambda$  é o vetor dos Multiplicadores de Lagrange

Derivando a função  $S$  em relação aos parâmetros  $\beta$  e  $\lambda$ ,

$$\partial S / \partial \beta = -2X'V^{-1}Y + 2(X'V^{-1}X)\beta + 2R'\lambda$$

$$\partial S / \partial \lambda = 2R\beta$$

Tomando-se as derivadas iguais a zero, constroem-se o seguinte sistema de equações, denominado Sistema de Equações Normais,

$$\begin{cases} -X'V^{-1}Y + (X'V^{-1}X)\hat{\beta}^{\#} + R'\hat{\lambda}^{\#} = 0 & (4.10) \\ R\hat{\beta}^{\#} = 0 & (4.11) \end{cases}$$

Matricialmente, pode-se representa-lo como,

$$\begin{bmatrix} X'V^{-1}X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta}^{\#} \\ \hat{\lambda}^{\#} \end{bmatrix} = \begin{bmatrix} X'V^{-1}Y \\ 0 \end{bmatrix}$$

Pré-multiplicando a equação (4.10) por  $R(X'V^{-1}X)^{-1}$ ,  
obtem-se:

$$\begin{aligned} -R(X'V^{-1}X)^{-1}X'V^{-1}Y + R(X'V^{-1}X)^{-1}(X'V^{-1}X)\hat{\beta}^{\#} + R(X'V^{-1}X)^{-1}R'\hat{\lambda}^{\#} &= 0 \\ -R(X'V^{-1}X)^{-1}X'V^{-1}Y + \hat{\beta}^{\#} + R(X'V^{-1}X)^{-1}R'\hat{\lambda}^{\#} &= 0 \end{aligned}$$

como  $(X'V^{-1}X)^{-1}X'V^{-1}Y = \hat{\beta}^*$ , no modelo generalizado irrestrito (4.4), tem-se que:

$$-\hat{\beta}^* + \hat{\beta}^{\#} + R(X'V^{-1}X)^{-1}R'\hat{\lambda}^{\#} = 0$$

Substituindo esse resultado em (4.11) e notando que  $R(X'V^{-1}X)^{-1}R'$  é não singular, tem-se:

$$\begin{aligned} R(X'V^{-1}X)^{-1}R'\hat{\lambda}^{\#} &= R\hat{\beta}^*, \text{ e portanto} \\ \hat{\lambda}^{\#} &= [R(X'V^{-1}X)^{-1}R']^{-1}R\hat{\beta}^*, \text{ que, substituindo-se em (4.10),} \\ \text{resulta:} \end{aligned}$$

$$\begin{aligned} -X'V^{-1}X + (X'V^{-1}Y)\hat{\beta}^{\#} + R'[R(X'V^{-1}X)^{-1}R']^{-1}R\hat{\beta}^* &= 0 \\ (X'V^{-1}Y)\hat{\beta}^{\#} &= X'V^{-1}Y - R'[R(X'V^{-1}X)^{-1}R']^{-1}R\hat{\beta}^* \end{aligned}$$

Pré-multiplicando por  $(X'V^{-1}X)^{-1}$ , obtem-se:

$$\begin{aligned} \hat{\beta}^{\#} &= (X'V^{-1}Y)X'V^{-1}Y - (X'V^{-1}Y)^{-1}R'[R(X'V^{-1}X)^{-1}R']^{-1}R\hat{\beta}^* \\ \hat{\beta}^{\#} &= \hat{\beta}^* - (X'V^{-1}Y)^{-1}R'[R(X'V^{-1}X)^{-1}R']^{-1}R\hat{\beta}^* \end{aligned} \quad (4.12)$$



Portanto, o estimador de mínimos quadrados generalizados restrito é igual ao estimador de mínimos quadrados generalizados irrestrito; mais uma combinação linear das restrições  $R\hat{\beta}^*$ .

A estimativa da soma de quadrados de resíduos para o modelo restrito é dada por:

$$\begin{aligned}\hat{SQR}^{\#} &= (Y - X\hat{\beta}^{\#})'V^{-1}(Y - X\hat{\beta}^{\#}) \\ &= Y'V^{-1}Y - 2Y'V^{-1}X\hat{\beta}^{\#} + \hat{\beta}^{\#}'X'V^{-1}X\hat{\beta}^{\#}\end{aligned}$$

Substituindo  $\hat{\beta}^{\#}$  por sua expressão em (4.12), tem-se:

$$\begin{aligned}\hat{SQR}^{\#} &= Y'V^{-1}Y - 2Y'V^{-1}X \langle \hat{\beta}^* - (X'V^{-1}Y)^{-1}R'[RCX'V^{-1}X]^{-1}R\hat{\beta}^* \rangle + \\ &\quad + \langle \hat{\beta}^* - (X'V^{-1}Y)^{-1}R'[RCX'V^{-1}X]^{-1}R\hat{\beta}^* \rangle' \langle X'V^{-1}X \rangle \cdot \\ &\quad \cdot \langle \hat{\beta}^* - (X'V^{-1}Y)^{-1}R'[RCX'V^{-1}X]^{-1}R\hat{\beta}^* \rangle\end{aligned}$$

que depois de algumas simplificações, resulta em:

$$\hat{SQR}^{\#} = Y'V^{-1}Y - 2Y'V^{-1}X\hat{\beta}^* + \hat{\beta}^*'X'V^{-1}X\hat{\beta}^* + \hat{\beta}^*'R'[RCX'V^{-1}X]^{-1}R\hat{\beta}^*$$

e finalmente,

$$\hat{SQR}^{\#} = \hat{SQR}^* + \hat{\beta}^*'R'[RCX'V^{-1}X]^{-1}R\hat{\beta}^* \quad (4.13)$$

onde,

$$\begin{aligned}\hat{SQR}^* &= Y'V^{-1}Y - 2Y'V^{-1}X\hat{\beta}^* + \hat{\beta}^*'X'V^{-1}X\hat{\beta}^* \\ &= (Y - X\hat{\beta}^*)'V^{-1}(Y - X\hat{\beta}^*)\end{aligned}$$

que é a soma de quadrados de resíduos do modelo generalizado irrestrito.

Tal como para os estimadores irrestritos de mínimos quadrados generalizados, consegue-se provar propriedades ótimas para os estimadores  $\hat{\beta}^{\#}$  e  $\hat{SQR}^{\#}$ .

## CAPÍTULO V

### O MODELO ESTATÍSTICO NA RESOLUÇÃO DO PROBLEMA BIOLÓGICO

#### V.1. INTRODUÇÃO

Nesta etapa do trabalho, busca-se adaptar a metodologia estatística proposta nos capítulos III e IV, ao problema biológico descrito no capítulo II, na tentativa de se obter estimativas para o parâmetro fisiológico Limiar de Anaerobiose (LA).

Um estudo detalhado do tema biológico abordado conduz à proposta de representação estatística do fenômeno, através de um modelo de regressão linear bi-segmentado, restrito nos parâmetros e sujeito a resíduos autocorrelacionados, sendo que assim, o ponto de junção de fases representaria o instante da investida do metabolismo anaeróbico no processo orgânico de obtenção de energia assim como proposto por Wasserman (1964).

A proposta de se estudar o fenômeno através de um modelo de regressão bi-segmentado baseia-se na tentativa de se formalizar estatisticamente os critérios de detecção do LA, oriundos de estudos fisiológicos (seção II.2.) e aplicados subjetivamente a respostas cardiorespiratórias não invasivas, provenientes de baterias de exercícios físicos de intensidade progressiva.

Revisando o que fora exposto na seção II.2., destacam-se alguns argumentos contundentes para a defesa do modelo estatístico proposto. Relembrando-se então que todos os critérios de detecção do LA citados admitiam uma mudança funcional no comportamento gráfico temporal de certas respostas cardiorespiratórias e utilizavam-se deste ponto de alteração para quantificarem o seu valor. Desta forma, estatisticamente, raciocina-se em termos de um relacionamento funcional entre variáveis, composto por dois sub-modelos distintos, de maneira a adotar o ponto de intercepto entre tais sub-modelos como um indicador no cálculo das estimativas do LA.

Consta ainda da proposta que tais sub-modelos deveriam ser lineares nos parâmetros o que, inicialmente, poderia ser justificado através do argumento da facilidade computacional no seu trato, não havendo contudo, prejuízo à estimação do parâmetro de interesse.

Visando adequar o modelo estatístico proposto ao fenômeno observado, algumas restrições limitantes a estimação dos parâmetros do modelo são impostas. Considera-se portanto, um modelo contínuo no instante da transição das fases, visto que fisiologicamente, a percepção sobre a investida do metabolismo energético se dá de forma progressiva porém branda.

Por razão da disponibilidade das informações requeridas, o modelo proposto será ajustado através das observações relativas a resposta Ventilação Minuto Respiratória (Volume de ar inspirado por minuto) através dos intervalos de Tempo de duração de um ciclo respiratório, decorrentes da experimentação realizada em alguns indivíduos sadios, submetidos a esforço físico progressivo (ver capítulo VI). Esta escolha pode também ser justificada através de prévios resultados obtidos por Martins (1986), que a adotou, por questão de conveniência metodológica, visto que, o ponto de inflexão na curva da resposta Ventilação seria, visualmente, mais acentuado do que o de outras respostas observadas.

De certa forma, a característica temporal das observações a serem submetidas à análise responde pela presença da hipótese de autocorrelação serial nos resíduos do modelo proposto e pela consequente utilização da metodologia dos mínimos quadrados generalizados restritos, na obtenção dos resultados requeridos.

Ainda assim, uma análise preliminar informal sobre os resíduos estimados do modelo adotado, através de mínimos quadrados ordinários, indicou a presença de comportamentos autorregressivos de ordem 1 e de ordem 2, sobre diferentes conjuntos de observações analisados. Desta forma, visto a necessidade de se propor um procedimento padronizado de análise e ainda, devido a facilidade no trato com resíduos autorregressivos de ordem 1, através de mínimos quadrados generalizados, convencionou-se adotar tal comportamento ao modelo proposto (seção V.2.).

Levando em consideração os padrões adotados por Wasserman na estimação visual do LA (seção II.2), uma análise preliminar do comportamento gráfico das curvas da resposta Ventilação versus Tempo, ao longo de algumas sessões de exercício físico progressivo, culminou na tentativa do ajuste de dois modelos bi-segmentados diferentes:

1. MODELO LL : composto por uma curva polinomial de grau um (reta) na primeira fase, seguida por outra curva polinomial de grau um (reta) na segunda fase;
2. MODELO LQ : composto por uma curva polinomial de grau um (reta) na primeira fase, seguida de uma curva polinomial de grau dois (parábola) na segunda fase.

Neste momento, discute-se a validade do carácter formal da metodologia proposta, visto que os resultados obtidos através de determinado modelo, podem diferir muito dos resultados

obtidos através de um outro modelo qualquer adotado. Desta forma, o resultado então acolhido estaria embutido de certo carácter subjetivo, o que, a princípio, foi o que se desejou evitar.

Comenta-se que, estatisticamente, tal controvérsia poderia ser facilmente resolvida mediante a aplicação de testes de hipóteses sobre a redução na soma de quadrados de resíduos entre um e outro modelo, porém, tais inferências, como fôra anteriormente falado, fogem do escopo deste trabalho e não serão analisadas.

Assim, antecipadamente, adota-se para efeito de resultados acolhidos, aqueles gerados através do MODELO LQ composto por um segmento de reta seguido de um segmento curvo em forma de parábola (RETA-PARABOLA), visto que, tal formulação permite uma representação mais próxima dos critérios fisiológicos de detecção discutidos. Os resultados obtidos através do MODELO LL (RETA-RETA), serão apresentados sob a forma de ilustração, de forma a gerar argumentos quando na comparação com os resultados visuais descritos por Martins (1986) (sessão VI.2.).

Na seção seguinte, formulam-se estatisticamente as idéias expostas até então, culminando com a apresentação do procedimento de estimação adotado.

Na seção V.3. apresenta-se um estudo ilustrativo acerca do comportamento da metodologia proposta, através de conjuntos de observações simuladas.

## V.2. FORMULAÇÃO DO PROBLEMA

Descrevendo o Modelo LQ comentado, sob a forma notacional, tem-se:

$$y_t = \begin{cases} \theta_{11} + \theta_{12} x_t + v_t & t = 1, 2, \dots, I \\ \theta_{21} + \theta_{22} x_t + \theta_{23} x_t^2 + v_t & t = I+1, I+2, \dots, N \end{cases} \quad (5.1)$$

$$\text{com } v_t = \rho v_{t-1} + \varepsilon_t$$

sujeito à restrição de continuidade em  $X = c$ , dada por:

$$\theta_{11} + \theta_{12} c = \theta_{21} + \theta_{22} c + \theta_{23} c^2$$

onde,  $Y$  : Ventilação minuto respiratória,

$X$  : Tempo de duração de um ciclo respiratório,

$$E(v) = 0, \quad E(\varepsilon) = 0,$$

$$V(v) = V\sigma^2, \quad V(\varepsilon) = I\sigma^2,$$

$$|\rho| < 1$$

Através do método dos mínimos quadrados generalizados (ver seção IV.3), o modelo poderá ser expresso, matricialmente, da seguinte forma:



$$PY = PX\theta + Pv$$

$$\text{sujeito a } R\theta = 0$$

onde,

$$Y = (Y1 \mid Y2)' = (y_1 \ y_2 \ \dots \ y_I \mid y_{I+1} \ y_{I+2} \ \dots \ y_N)'$$

$$v = (v1 \mid v2)' = (v_1 \ v_2 \ \dots \ v_I \mid v_{I+1} \ v_{I+2} \ \dots \ v_N)'$$

$$\theta = (\theta1 \mid \theta2)' = (\theta_{11} \ \theta_{12} \mid \theta_{21} \ \theta_{22} \ \theta_{23})'$$

$$X = \left( \begin{array}{c|c} X1 & 0 \\ \hline 0 & X2 \end{array} \right) = \left( \begin{array}{c|ccc} 1 & x_1 & & \\ \dots & & & 0 \\ 1 & x_I & & \\ \hline & & 1 & x_{I+1} & x_{I+1}^2 \\ 0 & & & \dots & \\ & & 1 & x_N & x_N^2 \end{array} \right)$$

$$R = (1 \ c \ -1 \ -c \ -c^2)$$

$$V = (1-\rho^2)^{-1} \left[ \begin{array}{c|c} v_1 & 0 \\ \hline 0 & v_2 \end{array} \right]$$

onde,

$$v_1 = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{I-1} \\ \rho & 1 & \rho & \dots & \rho^{I-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{I-3} \\ & & & \dots & \\ \rho^{I-1} & \rho^{I-2} & \rho^{I-3} & \dots & 1 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-I-1} \\ \rho & 1 & \rho & \dots & \rho^{N-I-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-I-3} \\ & & & \dots & \\ \rho^{N-I-1} & \rho^{N-I-2} & \rho^{N-I-3} & \dots & 1 \end{bmatrix}$$

e ainda,

$$V^{-1} = \left[ \begin{array}{c|c} V_1^{-1} & 0 \\ \hline 0 & V_2^{-1} \end{array} \right]$$

$$= \left[ \begin{array}{cccccc|cccc} & & & & \text{col I} & & & & & \\ 1 & -\rho & 0 & \dots & 0 & 0 & | & & & & \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 & | & & & & \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 & | & & & 0 & \\ & & & \dots & & & | & & & & \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho & | & & & & \\ 0 & 0 & 0 & \dots & -\rho & 1 & | & & & & \text{lin I} \\ \hline & & & & & & | & 1 & -\rho & 0 & \dots & 0 & 0 & \text{lin I+1} \\ & & & & & & | & -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 & \\ & & 0 & & & & | & 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 & \\ & & & & & & | & & & \dots & & & & \\ & & & & & & | & 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho & \\ & & & & & & | & 0 & 0 & 0 & \dots & -\rho & 1 & \end{array} \right]$$

e,

$$P = \left[ \begin{array}{c|c} P1 & 0 \\ \hline 0 & P2 \end{array} \right] = \left[ \begin{array}{cccccc|cccc} \sqrt{1-\rho^2} & 0 & \dots & 0 & 0 & 0 & & & & \\ -\rho & 1 & \dots & 0 & 0 & 0 & & & & \\ & & \dots & & & & & 0 & & \\ 0 & 0 & \dots & -\rho & 1 & 0 & & & & \\ \hline & & & & & & \sqrt{1-\rho^2} & 0 & \dots & 0 & 0 \\ & & & & & & -\rho & 1 & \dots & 0 & 0 \\ & & 0 & & & & & & \dots & & \\ & & & & & & 0 & 0 & \dots & -\rho & 1 \end{array} \right]$$

De forma que,

$$\hat{\theta}^{\#} = \hat{\theta}^* - (X'V^{-1}X)^{-1}R'[R(X'V^{-1}X)^{-1}R']^{-1}R\hat{\theta}^* \quad (5.2)$$

$$\hat{SSQR}^{\#} = \hat{SSQR}^* + \hat{\theta}^{*'}R'[R(X'V^{-1}X)^{-1}R']^{-1}R\hat{\theta}^* \quad (5.3)$$

onde,

$$\hat{\theta}^* = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad \text{e}$$

$$\hat{SSQR}^* = (Y - X\hat{\theta}^*)'V^{-1}(Y - X\hat{\theta}^*)$$

Formalizado notacionalmente o modelo proposto e definido o método dos mínimos quadrados generalizados para o ajuste dos parâmetros  $\theta$ , discute-se na seção seguinte, o procedimento de conduta adotado na estimação dos parâmetros de real interesse,  $c$  e  $I$ .

## V.2.1. PROCEDIMENTO DE ESTIMAÇÃO DO PONTO DE JUNÇÃO

Recorda-se, quando na seção III.3., sobre os procedimentos de estimação dos parâmetros de junção de um modelo bi-segmentado, onde verificou-se uma certa flexibilidade metodológica no procedimento "alternativo" de Lerman (1980).

Retomando então o estudo do problema biológico, observa-se que existem razões fisiológicas contundentes para se afirmar que o parâmetro de interesse pesquisado, LA, certamente ocorre em um intervalo de Tempo previsível, correspondente a certos níveis de esforço físico dispendidos distantes da exaustão, o que, estatisticamente, delimita regiões no espaço da variável Tempo nas quais seria inviável a pesquisa pela sua estimação, ou seja, por razões fisiológicas, ao parâmetro LA, poderia-se atribuir uma certa distribuição de probabilidades conhecida. Contudo, reforça-se a posição de não se fazer uso de argumentos Bayesianos na análise do problema abordado.

Além disso, um prosseguimento inevitável dessa pesquisa aponta para o ajuste de modelos de regressão bi-segmentados, compostos por sub-modelos não lineares, tendo em vista uma melhor adaptação do modelo estatístico proposto aos critérios fisiológicos de detecção do LA descritos.

Desta forma, o problema biológico tratado, reflete a conveniência do uso do procedimento "alternativo" de Lerman em sua resolução.

A seguir, descreve-se portanto, tal maneira de se abordar o problema. Caracteriza-se a sequência de passos a serem descritas como um procedimento de busca, computacionalmente intensivo.

Passo 1. Fixado um  $\hat{I}$  ( $\hat{I} = 2, 3, \dots, N-3$ ), de forma que exista a possibilidade de ocorrer um ponto de junção no intervalo  $[x_{\hat{I}}^{\wedge}, x_{\hat{I}+1}^{\wedge}]$ , determina-se  $k$ , um divisor da amplitude  $(x_{\hat{I}+1}^{\wedge} - x_{\hat{I}}^{\wedge})$ , de acordo com a precisão com que se deseja estimar  $c$ . Calcula-se então:

$$d = (x_{\hat{I}+1}^{\wedge} - x_{\hat{I}}^{\wedge}) / k$$

Passo 2. Ao limite inferior do intervalo tomado,  $x_{\hat{I}}^{\wedge}$ , adiciona-se o valor do incremento  $d$  calculado no passo 1, de forma a obter  $(k-1)$  valores intermediários no intervalo  $[x_{\hat{I}}^{\wedge}, x_{\hat{I}+1}^{\wedge}]$ , ou seja:

$$x_{\hat{I}}^{\wedge} < x_{\hat{I}}^{\wedge} + d < x_{\hat{I}}^{\wedge} + 2d < \dots < x_{\hat{I}}^{\wedge} + (k-1)d < x_{\hat{I}+1}^{\wedge}$$

Passo 3. Para cada valor obtido no passo anterior (possíveis candidatos a pontos de junção) procede-se da seguinte forma:

- Estima-se  $\theta$ , através de mínimos quadrados ordinários restritos, tal como apresentado no apêndice 2:

$$\hat{\theta}^+ = \hat{\theta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\hat{\theta}$$

$$\text{onde } \hat{\theta} = (X'X)^{-1}X'Y;$$

- Calculam-se os resíduos estimados desse modelo,

$$\hat{v}^+ = Y - X\hat{\theta}^+;$$

- Calcula-se o valor de  $\hat{\rho}$  tal como descrito no procedimento de Orcutt-Cochrane (sessão IV.2),

$$\hat{\rho} = \left( \sum_{t=2}^N \hat{v}_t^+ \hat{v}_{t-1}^+ \right) / \sum_{t=1}^N \hat{v}_t^{+2}$$

- Utiliza-se o valor da estimativa de  $\hat{\rho}$ , calculada no passo anterior, para construir a matriz P, e portanto, as matrizes V e  $V^{-1}$ , da forma como estão apresentadas na sessão V.2.;
- Estima-se, através de mínimos quadrados generalizados, os parâmetros  $\theta$ , através de  $\hat{\theta}^{\#}$  (4.12) e então, o valor da estimativa  $\hat{SSQR}^{\#}$  (4.13), que deve ser correspondentemente anotada;

Passo 4. Após repetir os passos 1 a 3, para todos os possíveis  $\hat{I}$ 's viáveis, toma-se como estimativa do parâmetro c, aquele

valor fixado como ponto de junção, correspondente ao menor valor de  $\hat{SSQR}^{\#}$ , obtido sobre todos os outros calculados.

Um algoritmo computacional foi desenvolvido em linguagem FORTRAN VII, especificamente para a solução deste problema. Nele constam rotinas de manipulação e cálculos matriciais tal qual necessários à obtenção dos resultados esperados. O equipamento necessário para a execução desta implementação, trata-se simplesmente de um micro computador compatível com o PC-IBM.

O processamento dos dados foi realizado no Laboratório de Estatística Aplicada (LEA) do Instituto de Planejamento e Estudos Ambientais IPEA-UNESP, campus de Presidente Prudente (SP). As figuras que constam dos resultados obtidos foram construídas através de "softwares" gráficos disponíveis.

### V.3. ILUSTRAÇÃO - UM ESTUDO DE SIMULAÇÃO

Como fôra dito anteriormente, para fins de ilustração, a metodologia proposta (seção V.2.), foi utilizada no tratamento de alguns conjuntos de observações simuladas, e os respectivos resultados, colocados em discussão.

Assim, foram gerados três grupos (1,2 e 3) de observações de tamanho  $N = 40$ , relativos a componente determinística do modelo proposto (5.1), da seguinte forma: Para cada um dos três



grupos, estabeleceu-se determinados valores para os parâmetros  $\theta$  e para o parâmetro  $c$  que, propositalmente, fôra fixado no início, no meio e no fim do espaço de variação da variável Tempo estabelecida (Tempo  $\in [0,200]$ ). Assim, a componente determinística das observações simuladas, que constam da Tabela V.1., foram geradas a partir dos seguintes valores de parâmetros:

Grupo 1:  $\theta_{11} = 10.0$  ;  $\theta_{12} = 1.0$  ;  
 $\theta_{21} = 85.0$  ;  $\theta_{22} = -1.0$  ;  $\theta_{23} = 0.01$  ;  
 Intercepto:  $c = 50.0$

Grupo 2:  $\theta_{11} = 10.0$  ;  $\theta_{12} = 0.5$  ;  
 $\theta_{21} = 160.0$  ;  $\theta_{22} = -2.0$  ;  $\theta_{23} = 0.01$  ;  
 Intercepto:  $c = 100.0$

Grupo 3:  $\theta_{11} = 10.0$  ;  $\theta_{12} = 0.4$  ;  
 $\theta_{21} = 295.0$  ;  $\theta_{22} = -3.0$  ;  $\theta_{23} = 0.01$  ;  
 Intercepto:  $c = 150.0$

Logo após, obteve-se um vetor de tamanho  $N = 40$ , de valores gerados aleatória e independentemente, segundo a distribuição Normal de probabilidades, com parâmetros  $\mu = 0$  e  $\sigma = 4$ , isto é,  $\varepsilon_t \sim \text{iid NORMAL}(0, \sigma = 4)$ ,  $t = 1, 2, \dots, 40$ . Deve-se notar que não consta do modelo proposto (5.1) qualquer alusão à distribuição de probabilidades do vetor  $\varepsilon$  que, neste instante, foi gerado como Normal, sem que isso pudesse oferecer prejuízo a integridade do modelo.

Tomando por base esses valores, a componente aleatória do modelo fôra obtida através de duas transformações Markovianas, que caracterizavam a presença do parâmetro de autocorrelação serial, sob a forma de um distúrbio autorregressivo de ordem 1, ou seja,

$$v_t = \rho v_{t-1} + \epsilon_t.$$

Assim, dois novos grupos (A e B) de observações geradas (constam da Tabela V.1.) foram obtidos, na representação da componente estocástica do modelo, diferindo entre si, através do valor do parâmetro autorregressivo  $\rho$ , que propositalmente fôra fixado de forma a caracterizar a presença duvidosa ou a presença marcante do fator de autocorrelação, ou seja:

Grupo A :  $\rho = 0.6$  (presença duvidosa de autocorrelação);

Grupo B :  $\rho = 0.95$  (presença marcante de autocorrelação).

Finalizando, obtiveram-se seis grupos de valores simulados, a partir da composição aditiva das três componentes determinísticas (1,2 e 3) com as duas componentes estocásticas (A e B) geradas. Por convenção, os grupos de valores em estudo foram denominados SIMULA 1A, SIMULA 1B, SIMULA 2A, SIMULA 2B, SIMULA 3A e SIMULA 3B. Tais valores estão apresentados na Tabela V.2., e seus gráficos de dispersão constam das Figuras V.1, V.3, V.5, V.7, V.9 e V.11 respectivamente.

A cada um dos conjuntos de observações simulados, foi aplicado o procedimento de estimação descrito na seção anterior,

Tabela V.3, que resume as informações acerca do comportamento numérico mínimo das estimativas SSQR<sup>#</sup>, calculadas e apresentadas graficamente nas Figuras V.2, V.4, V.6, V.8, V.10 e V.12, respectivamente.

TABELA V.3. Resultados Do Estudo De Simulação

PARAMETRO	SIMULA 1A		SIMULA 1B	
	VALOR REAL	ESTIMATIVA	VALOR REAL	ESTIMATIVA
$\rho$	0.60	0.4196	0.95	0.8135
$\theta_{11}$	10.00	9.1891	10.00	7.7525
$\theta_{12}$	1.00	1.0536	1.00	1.1303
$\theta_{21}$	85.00	83.7200	85.00	97.5270
$\theta_{22}$	-1.00	-1.0361	-1.00	-1.3310
$\theta_{23}$	0.01	0.0102	0.01	0.0110
c	50.00	46.00	50.00	46.00

cont.

PARAMETRO	SIMULA 2A		SIMULA 2B	
	VALOR REAL	ESTIMATIVA	VALOR REAL	ESTIMATIVA
$\rho$	0.60	0.3591	0.95	0.8177
$\theta_{11}$	10.00	11.1781	10.00	9.1923
$\theta_{12}$	0.50	0.4519	0.50	0.4171
$\theta_{21}$	160.00	106.7699	160.00	118.6879
$\theta_{22}$	-2.00	-1.3303	-2.00	-1.6504
$\theta_{23}$	0.01	0.0079	0.01	0.0090
c	100.00	88.00	100.00	83.00

cont.

PARÂMETRO	SIMULA 3A		SIMULA 3B	
	VALOR REAL	ESTIMATIVA	VALOR REAL	ESTIMATIVA
$\rho$	0.60	0.3517	0.95	0.8093
$\theta_{11}$	10.00	10.4232	10.00	10.6439
$\theta_{12}$	0.40	0.3701	0.40	0.2803
$\theta_{21}$	295.00	-21.1018	295.00	40.5967
$\theta_{22}$	-3.00	0.7251	-3.00	-0.2505
$\theta_{23}$	0.01	-0.0008	0.01	0.0022
c	150.00	132.00	150.00	134.00

### V.3.1. DISCUSSÕES

Apesar de o estudo em questão prestar-se apenas à ilustração e não corresponder a um verdadeiro estudo de simulação, no qual exigiria-se a replicação exaustiva do processo executado, sob diferentes condições estocásticas, admite-se discutir seus resultados sob a alegação de levantamento de hipóteses acerca do procedimento de análise proposto..

Desta forma, relativo aos resultados apresentados na Tabela V.3., percebe-se primeiramente que o parâmetro autorregressivo  $\rho$  esteve sempre sub-estimado, indiferentemente ao grupo de observações analisado. De volta à seção IV.3., recorda-se que o procedimento de Orcutt-Cochrane adotado, poderia vir a ser utilizado iterativamente, de forma a obter uma estimativa convergente do parâmetro, o que, supostamente,

contribuiria para a melhoria da sua precisão. Entretanto, comenta-se que a utilização iterada desse procedimento significaria um aumento substancial no tempo de processamento dos dados necessário, culminando na ineficiência computacional do procedimento proposto.

Ainda acerca dos resultados obtidos, percebe-se que as estimativas dos parâmetros  $\theta$ , relativas ao sub-modelo polinomial quadrático, diferem substancialmente dos verdadeiros valores conhecidos, à medida em que o intercepto  $X = c$  desloca-se para a direita, ou seja, à medida em que reduz-se o número de observações utilizadas na sua estimação.

Finalizando, observa-se que as estimativas da abscissa do intercepto  $c$ , parâmetro de real interesse na análise, mostraram-se sempre abaixo do verdadeiro valor conhecido, isto é, elas sempre o sub-estimaram.

Novamente, retoma-se o comentário de que os resultados obtidos e, conseqüentemente, as discussões provenientes destes resultados, baseiam-se apenas em uma única realização estocástica do experimento e não podem, portanto, serem admitidas estatisticamente. Seguem-se as Tabelas e Figuras mencionadas ao longo dessa seção.

TABELA V.1. Valores Simulados Das Componentes Determinísticas e Aleatórias Do Modelo Proposto.

TEMPO	GRUPO 1	GRUPO 2	GRUPO 3	GRUPO A	GRUPO B
2	12.00	11.00	10.80	- 3.80	- 3.80
7	17.00	13.50	12.80	3.35	1.99
12	22.00	16.00	14.80	2.13	2.01
17	27.00	18.50	16.80	3.34	3.97
22	32.00	21.00	18.80	3.74	5.51
27	37.00	23.50	20.80	0.69	3.67
32	42.00	26.00	22.80	0.22	3.30
37	47.00	28.50	24.80	- 0.12	2.88
42	52.00	31.00	26.80	- 2.12	0.59
47	57.00	33.50	28.80	2.42	4.25
52	60.04	36.00	30.80	0.49	2.10
57	60.49	38.50	32.80	- 0.75	1.53
62	61.44	41.00	34.80	0.58	2.49
67	62.89	43.50	36.80	1.31	3.32
72	64.84	46.00	38.80	0.74	3.12
77	67.29	48.50	40.80	- 5.02	- 2.51
82	70.24	51.00	42.80	- 11.66	- 11.03
87	73.69	53.50	44.80	- 6.04	- 9.52
92	77.64	56.00	46.80	- 2.87	- 8.30
97	82.09	58.50	48.80	- 4.64	- 10.80
102	87.04	60.04	50.80	- 3.19	- 10.66
107	92.49	60.49	52.80	- 1.99	- 10.21
112	98.44	61.44	54.80	- 3.42	- 11.92
117	104.89	62.89	56.80	- 3.80	- 13.07
122	111.84	64.84	58.80	- 7.12	- 17.26
127	119.29	67.29	60.80	- 10.12	- 22.25
132	127.24	70.24	62.80	- 5.17	- 20.83
137	135.69	73.69	64.80	3.59	- 12.74
142	144.64	77.64	66.80	- 4.51	- 18.76
147	154.09	82.09	68.80	1.03	14.10
152	164.04	87.04	70.04	5.08	- 8.93
157	174.49	92.49	70.49	0.65	- 10.89
162	185.44	98.44	71.44	2.02	- 8.71
167	196.89	104.89	72.89	- 2.32	- 11.81
172	208.84	111.84	74.84	4.17	- 5.66
177	221.29	119.29	77.29	- 1.01	- 8.88
182	234.24	127.24	80.24	5.88	- 1.95
187	247.69	135.69	83.69	- 4.70	- 10.08
192	261.64	144.64	87.64	- 0.76	- 7.52
197	276.09	154.09	92.09	- 3.82	- 10.50

TABELA V.2. Valores Simulados de Observações.

TEMPO	SI 1A	SI 1B	SI 2A	SI 2B	SI 3A	SI 3B
2	8.10	8.10	7.10	7.10	6.90	6.90
7	20.35	18.99	16.85	15.49	16.15	14.79
12	24.13	24.01	18.13	18.01	16.93	16.81
17	30.34	30.97	21.84	22.47	20.14	20.77
22	35.74	37.51	24.74	26.51	22.54	24.31
27	37.69	40.67	24.19	27.17	21.49	24.47
32	42.22	45.30	26.22	29.30	23.02	26.10
37	46.88	49.88	28.38	31.38	24.68	27.68
42	49.88	52.59	28.88	31.59	24.68	27.39
47	59.42	61.25	35.92	37.75	31.22	33.05
52	59.55	62.14	35.51	38.10	30.31	32.90
57	59.74	62.02	37.75	40.03	32.05	34.33
62	62.02	63.93	41.58	43.49	35.38	37.29
67	64.20	66.21	44.81	46.82	38.11	40.12
72	65.57	67.96	46.74	49.12	39.54	41.92
77	62.27	64.78	43.48	45.99	35.78	38.29
82	58.58	59.21	39.34	39.97	31.14	31.77
87	67.65	64.17	47.46	43.98	38.76	35.28
92	74.77	69.34	53.13	47.70	43.93	38.50
97	77.45	71.28	53.86	47.70	44.16	38.00
102	83.85	76.38	56.85	49.38	47.61	40.14
107	90.50	82.28	58.50	50.28	50.81	42.59
112	95.02	86.52	58.02	49.52	51.38	42.88
117	101.09	91.82	59.09	49.82	53.00	43.73
122	104.72	94.57	57.72	47.57	51.68	41.54
127	109.17	97.04	57.17	45.04	50.68	38.55
132	122.07	106.41	65.07	49.41	57.63	41.97
137	139.28	122.95	77.28	60.95	68.39	52.06
142	140.13	125.88	73.13	58.88	62.29	48.04
147	155.12	139.99	83.12	67.99	69.83	54.70
152	169.12	155.11	92.12	78.11	75.12	61.11
157	175.14	163.60	93.14	81.60	71.14	59.60
162	187.46	176.73	100.46	89.73	73.46	62.73
167	194.57	185.08	102.57	93.08	70.57	61.08
172	213.01	203.18	116.01	106.18	79.00	69.17
177	220.28	212.41	118.28	110.41	76.28	68.41
182	240.12	232.29	133.12	125.29	86.12	78.29
187	242.99	237.61	130.99	125.61	78.99	73.61
192	260.88	254.12	143.88	137.12	86.88	80.12
197	272.27	265.59	150.27	143.59	88.27	81.59

FIGURA V.1. Gráfico de dispersão dos valores gerados SIMULA 1A  
contra o TEMPO

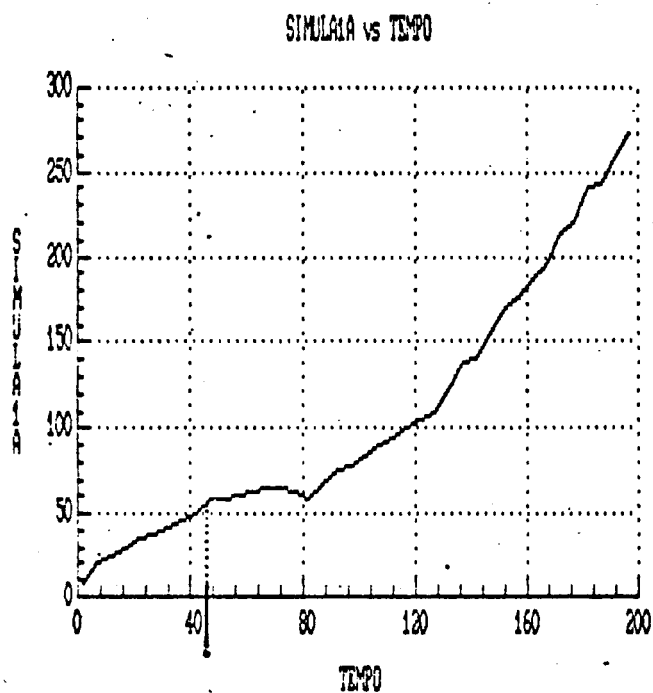


FIGURA V.2. Gráfico de dispersão dos valores estimados  
 $\hat{SSQR}^{\#} - 1A$  contra o TEMPO

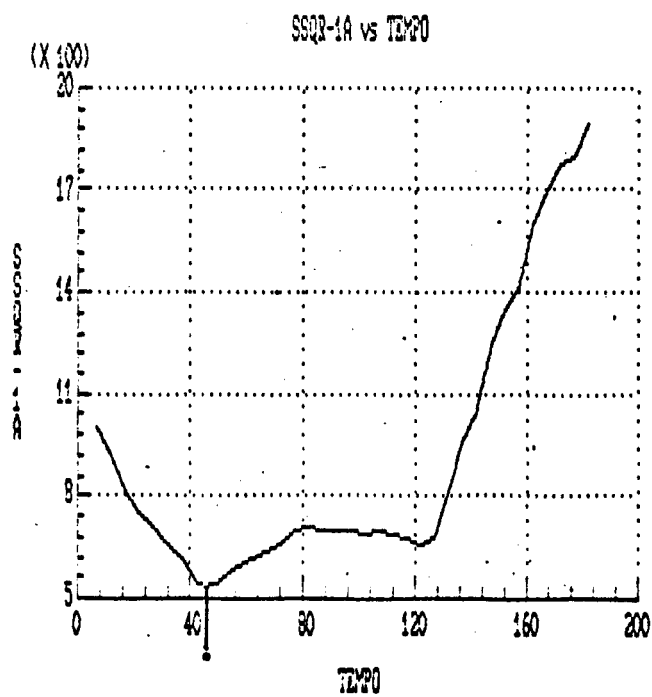




FIGURA V.3. Gráfico de dispersão dos valores gerados SIMULA 1B  
contra o TEMPO

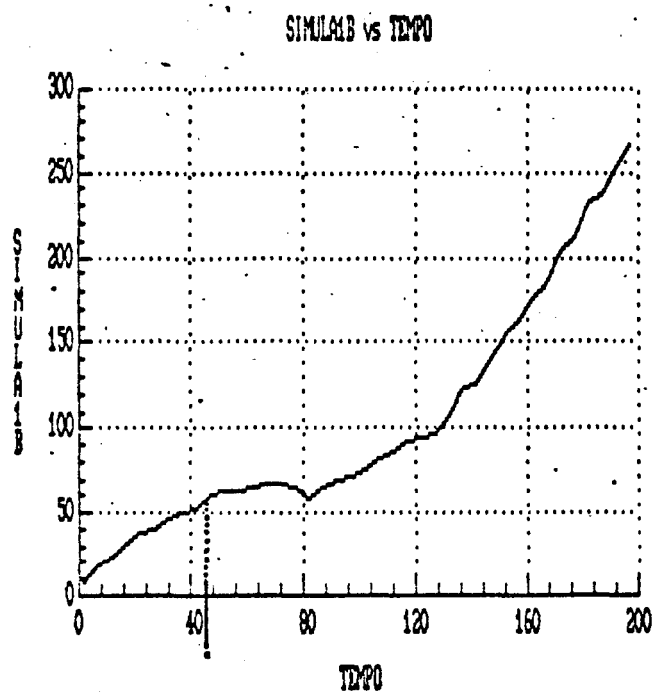


FIGURA V.4. Gráfico de dispersão dos valores estimados  
 $\hat{SSQR}^* - 1B$  contra o TEMPO

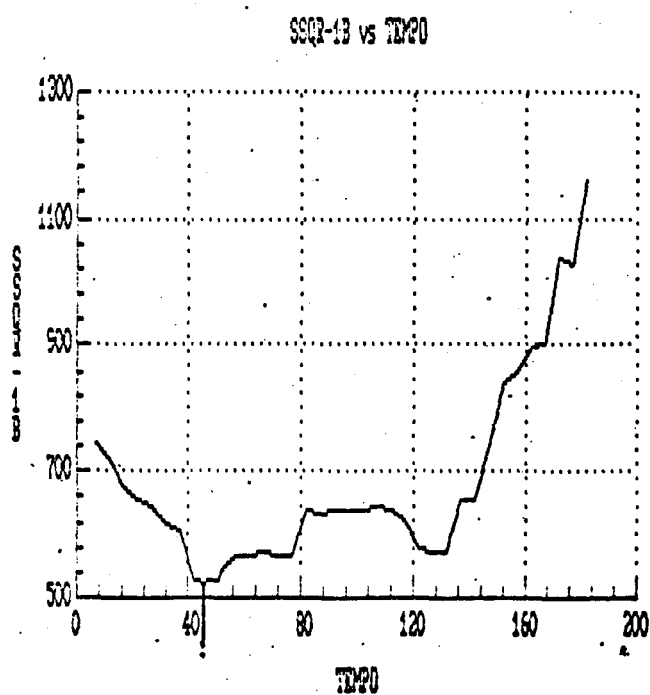


FIGURA V.5. Gráfico de dispersão dos valores gerados SIMULA 2A  
contra o TEMPO

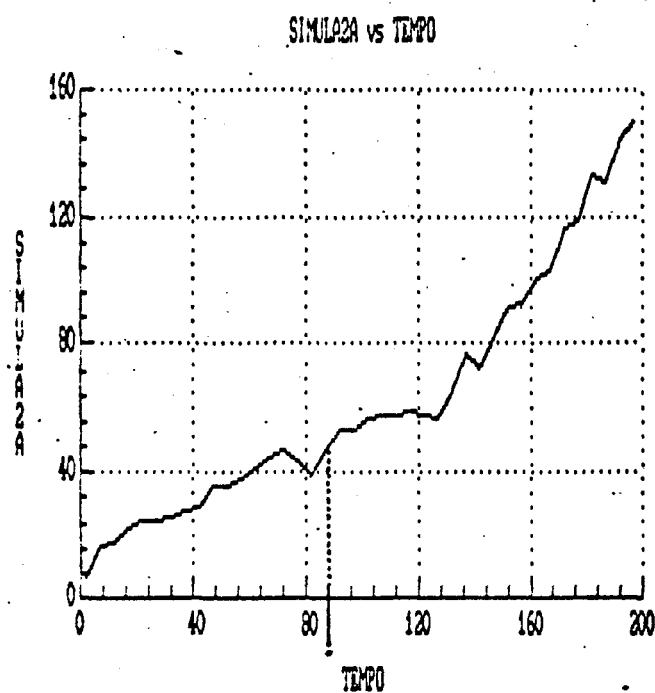


FIGURA V.6. Gráfico de dispersão dos valores estimados  
 $\hat{SSQR}^{\#} - 2A$  contra o TEMPO

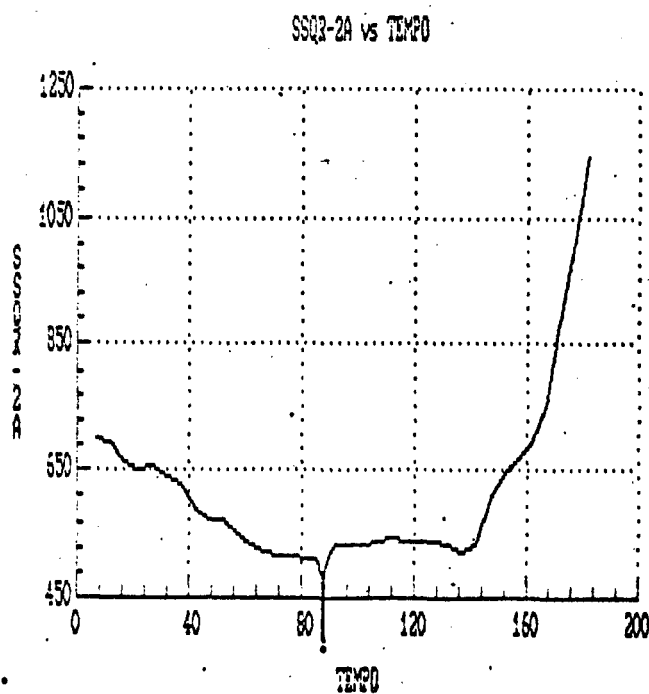


FIGURA V.7. Gráfico de dispersão dos valores gerados SIMULA 2B  
contra o TEMPO

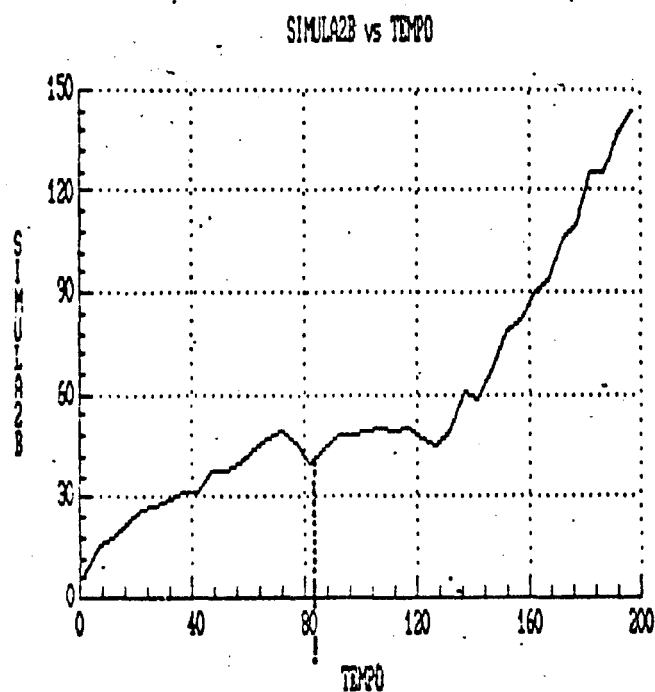


FIGURA V.8. Gráfico de dispersão dos valores estimados  
 $\hat{SSQR}^{\#}$ - 2B contra o TEMPO

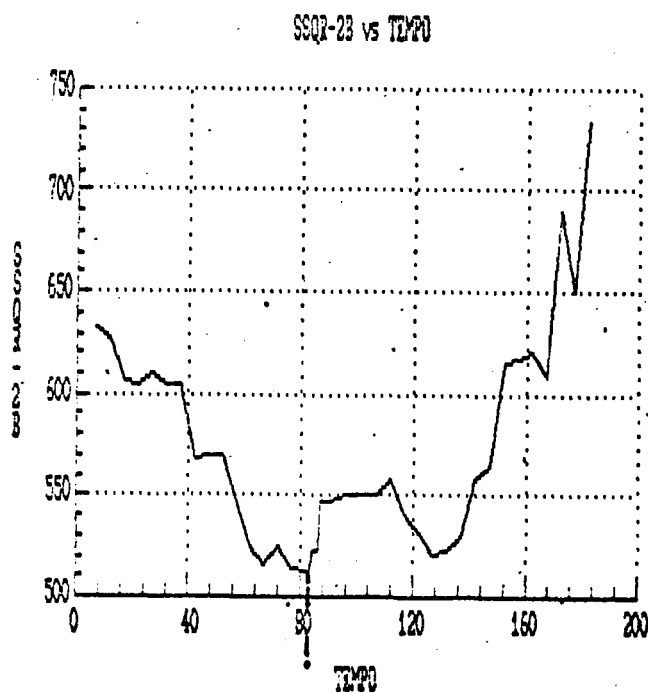


FIGURA V.9. Gráfico de dispersão dos valores gerados SIMULA 3A  
contra o TEMPO

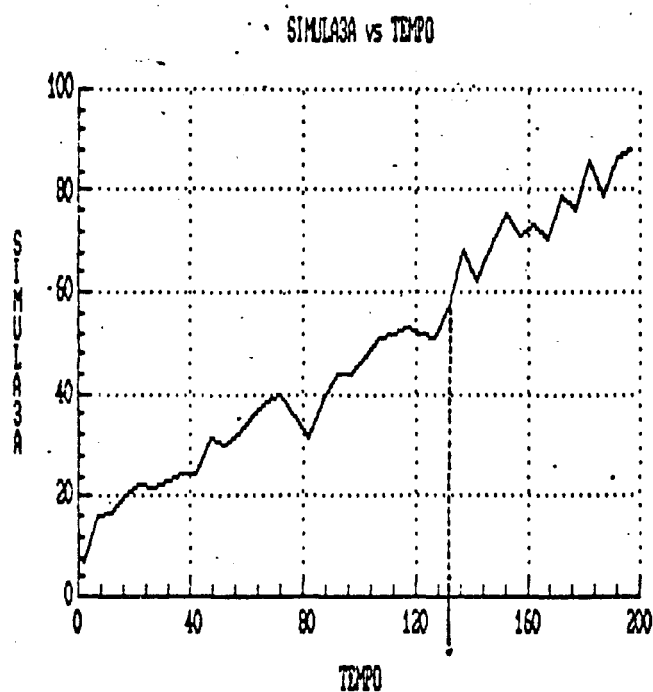


FIGURA V.10. Gráfico de dispersão dos valores estimados  
 $\hat{SSQR}^{\#} - 3A$  contra o TEMPO

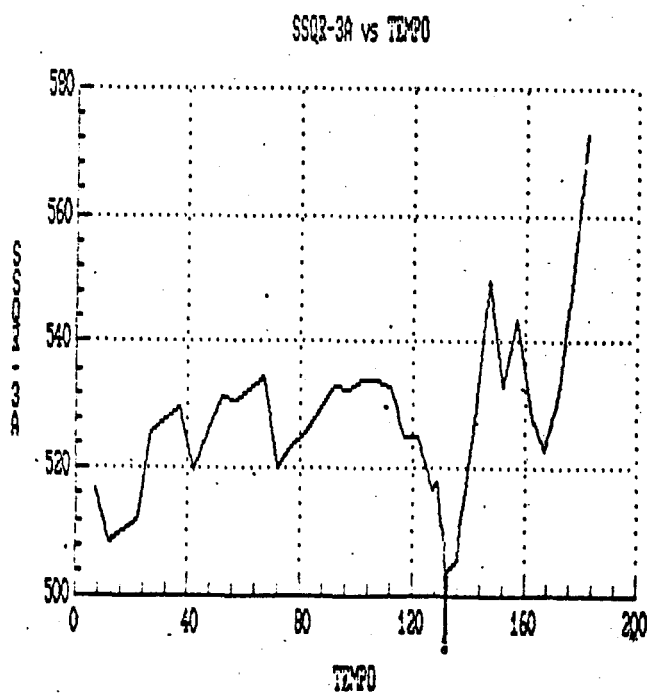


FIGURA V.11. Gráfico de dispersão dos valores gerados SIMULA 3B  
contra o TEMPO

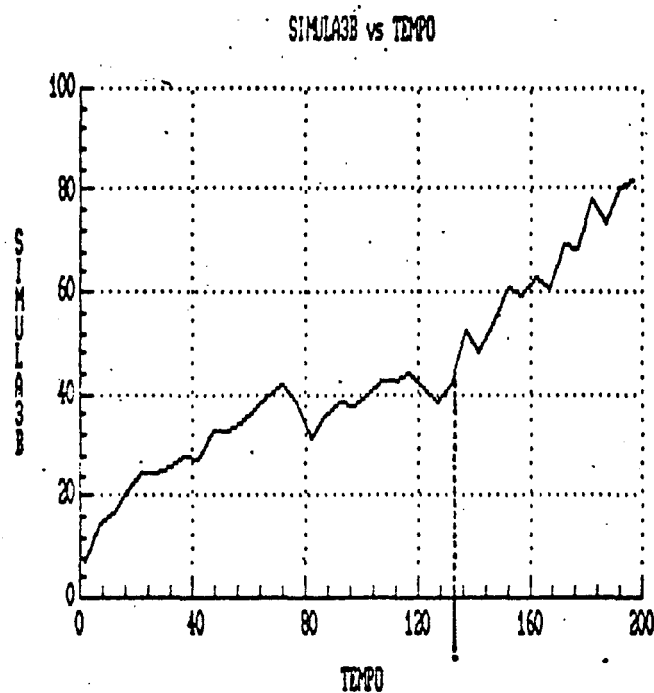
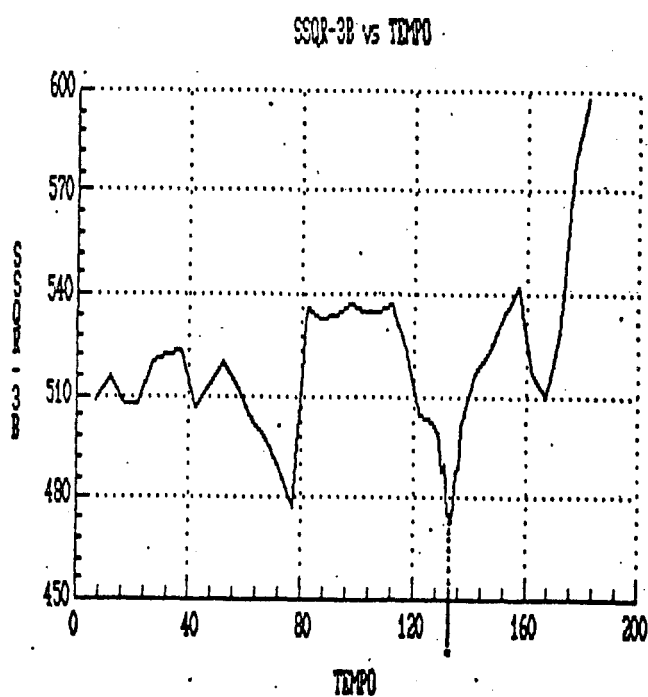


FIGURA V.12. Gráfico de dispersão dos valores estimados  $\hat{SSQR}^* - 3B$  contra o TEMPO



## CAPÍTULO VI

### APLICAÇÃO

#### VI.1. O DELINEAMENTO EXPERIMENTAL

Os resultados observados que são utilizados neste trabalho, foram coletados e analisados por Martins (1986), através de um protocolo experimental desenvolvido pela Seção de Hemodinâmica e Função Pulmonar do Hospital das Clínicas FMRP - USP, e que se assemelha aos protocolos experimentais discutidos em literaturas afins. Deste, consta uma série de equipamentos analógicos sofisticados, de tecnologia atual, utilizados na observação não invasiva de algumas respostas cardiorespiratórias de indivíduos submetidos a exercícios físicos progressivos.

Desta forma, destacam-se alguns elementos componentes deste protocolo. O esforço físico progressivo experimental, foi obtido através de uma bicicleta ergométrica especialmente

preparada. Os potenciais cardíacos foram medidos por um eletrocardiografo no qual um módulo amplificador, associado ao equipamento, computou a frequência cardíaca instantânea, a partir do reconhecimento do sinal de um QRS destacado e, um fluxômetro antecedendo a entrada de uma válvula de ar, registrou o volume corrente de ar inspirado.

O experimento foi realizado em cinco indivíduos do sexo masculino, não fumantes, que desenvolviam na época atividade física moderada. Cada indivíduo foi submetido a quatro sessões experimentais, as quais se realizaram obedecendo sempre ao mesmo período matinal ou vespertino, a fim de se evitar possíveis variações circadianas. Em cada uma das sessões experimentais variava-se a forma funcional do acréscimo do esforço físico requerido.

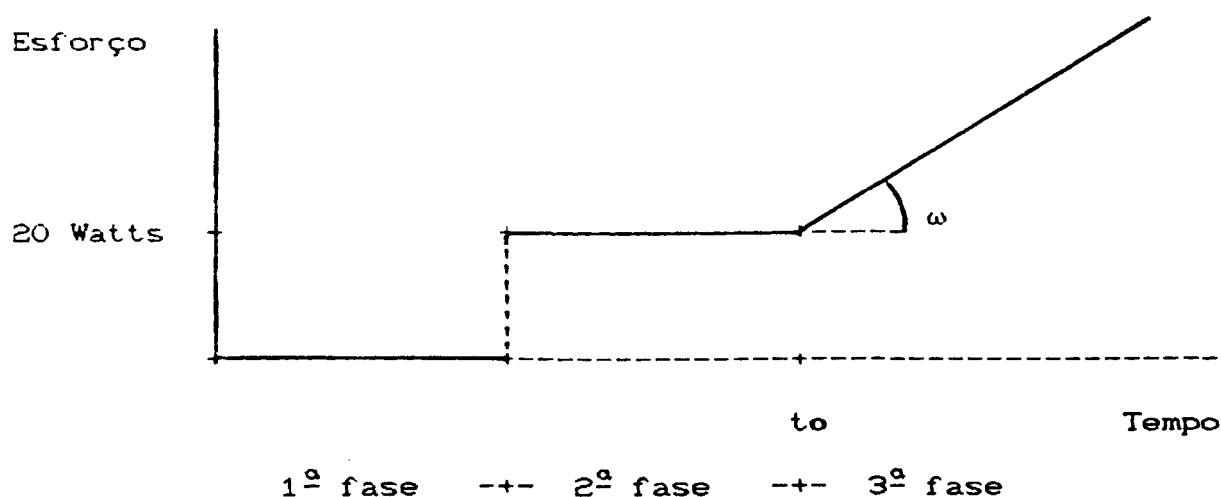
A temperatura ambiente do laboratório foi mantida artificialmente entre 20 e 22 graus Célcus, e todos foram orientados no sentido de não realizarem esforço físico com os membros superiores no apoio do guidão da bicicleta. O posicionamento da peça bucal e os ajustes do selim e do guidão foram feitos de modo a proporcionar uma posição confortável aos indivíduos durante as sessões experimentais de esforço.

Na primeira sessão experimental foram aplicados testes nos quais o esforço dispendido caracterizava a forma de uma rampa, cuja inclinação variava entre 25 e 108 Watts/minuto. A partir dos resultados obtidos nesta sessão, Martins (1986) quantificou

visualmente valores para o LA, através do ponto de inflexão da curva da resposta Ventilação contra o Tempo (ver Tabela VI.2.).

Durante a realização desta primeira sessão experimental, determinaram-se três fases de medições analógicas das respostas cardiorespiratórias. Na primeira fase as medições eram realizadas mantendo os indivíduos em repouso. Na segunda fase, o início do esforço se fez de forma constante e branda através de uma carga de 20 Watts no pedal da bicicleta durante um período de aproximadamente 5 minutos. Na terceira fase, ocorreu o início real da rampa (Tempo =  $t_0$ ), a partir do qual, a carga no pedal cresceu instantaneamente com o decorrer do tempo. A Figura VI.1, a seguir, ilustra esta descrição.

FIGURA VI.1. Padrão do esforço físico dispendido durante a primeira sessão experimental.





A segunda sessão experimental, consistiu de testes em forma de degraus e na terceira e quarta sessões, foram aplicados testes em formas senoidais..

As variáveis experimentais Carga da bicicleta, Velocidade de rotação do pedal, Eletrocardiograma, Volume corrente de ar inspirado e Frequência cardíaca instantânea, foram registradas analógica-digitalmente durante o desenrolar dos experimentos, com frequência de amostragem de 50 milisegundos.

Neste trabalho, fez-se uso dos resultados obtidos durante a terceira fase da primeira sessão experimental, relativos à resposta Ventilação minuto respiratória (Volume corrente de ar inspirado por minuto) em padrões de esforço na forma de rampas cujas inclinações variaram entre 25 e 108 Watts/minuto.

Dispõe-se, entretanto, apenas das observações relativas a quatro dos cinco indivíduos analisados por Martins (1986). Em correspondência ao trabalho de Martins (1986), estes serão denominados, a partir de agora, pelas letras T, U, X e Z. Assim, o modelo proposto para a estimação do LA (5.1) foi aplicado às observações experimentais da resposta Ventilação versus Tempo, dos quatro indivíduos citados, submetidos a esforço físico progressivo em rampas com as seguintes inclinações:

Indivíduo T : rampa de 28 W/min e  
rampa de 64 W/min;

Indivíduo U : rampa de 28 W/min e  
rampa de 47 W/min;

Indivíduo X : rampa de 28 W/min e  
rampa de 56 W/min;

Indivíduo Z : rampa de 32 W/min e  
rampa de 108 W/min.

A partir da estimação estatística do valor de Tempo onde ocorreu a mudança funcional na curva da resposta Ventilação, denotado por  $\hat{c}$ , o valor estimado do LA, em unidades de carga (Watts), foi calculado através da seguinte expressão:

$$\hat{LA} = 20 + \omega (\hat{c} - t_0) / 60 \quad (6.1)$$

resultante do padrão de esforço físico dispendido (ver Figura VI.1).

## VI.2. RESULTADOS

Constam do Apêndice 1 os valores das observações utilizadas nesta análise correspondentes aos indivíduos e as inclinações no padrão de esforço físico à qual foram submetidos. A dispersão de tais observações, podem ser vistas também sob a forma gráfica através das Figuras VI.2, VI.4, VI.6, VI.8, VI.10, VI.12, VI.14 e VI.16, respectivamente.

Cada um dos conjuntos de observações apresentados foi submetido ao procedimento de estimação proposto (seção V.2.), através de "software" especificamente construído. Para tanto, foram tomados  $\hat{I}$ 's que variaram desde 2 até N-3. Em relação a divisão do intervalo  $[x_{\hat{I}}, x_{\hat{I}+1}]$ , estabeleceu-se que para cada nova abscissa de Tempo fixada, deveria ocorrer uma diferença de aproximadamente 0.5 Watts de esforço físico dispendido. Desta forma, correspondente a cada inclinação no padrão do esforço dispendido durante as sessões experimentais, calculou-se um incremento  $d_{\omega}$  ( $\omega = 28, 32, 47, 56, 64$  e  $108$  Watts/minuto), que estão apresentados na Tabela VI.1.

TABELA VI.1. Incremento de Tempo relativo às inclinações no esforço físico dispendido.

$\omega$ (W/min)	$d_{\omega}$ (seg)
28	1.0
32	1.0
47	0.5
56	0.5
64	0.5
108	0.25

Portanto, o número de abscissas fixadas em cada conjunto de observações experimentais analisado, relativo a cada escolha de  $\hat{I}$ , foi dado por:

$$k = (x_{I+1}^{(0)} - x_I^{(0)}) / d_{\omega}$$

A Tabela VI.2 contém os resultados obtidos, na qual apresenta-se os valores estimados da abscissa Tempo, relativos às mudanças funcionais no comportamento da curva da Ventilação e o correspondente esforço físico instantâneo dispendido, calculado através da expressão (6.1), ou seja, o valor estimado do LA, tanto para os casos da estimação visual, obtidas por Martins (1986), quanto para os modelos ajustados LINEAR-LINEAR e LINEAR-QUADRÁTICO. Observa-se que em alguns casos, as estimativas analíticas são apresentadas sob a forma intervalar, visto não ter sido possível detectar diferenças numéricas nos valores de  $\hat{SSQR}^{\#}$  entre as abscissas intermediárias fixadas.

Tais resultados resumem portanto as informações contidas nas Figuras VI.3, VI.5, VI.7, VI.9, VI.11, VI.13, VI.15 e VI.17, relativas ao comportamento minimal das estimativas  $\hat{SSQR}^{\#}$  (5.3) calculadas.

Tendo em vista que as estimativas visuais quantificadas por Martins (1986) possuem um vício metodológico intencional, que resultou na sub-estimação dos parâmetros desejados, propõe-se no Apêndice 3 um estudo comparativo, comumente encontrado na

literatura bio-médica específica, entre os resultados obtidos através do procedimento proposto e alguns resultados visuais quantificados por fisiologistas e matemáticos numéricos.

TABELA VI.2. Resultados comparativos das estimativas do LA através do critério visual e do procedimento proposto.

		VISUAL		MODELO LL		MODELO LQ	
		$\hat{c}$	$\hat{LA}$	$\hat{c}$	$\hat{LA}$	$\hat{c}$	$\hat{LA}$
		(seg)	(w)	(seg)	(w)	(seg)	(w)
T	28	133	82	158.40	93.92	85.15	59.74
	64	70	95	91.79-92.79	117.91-118.98	16.60-17.15	37.71-38.29
U	28	113	72	133.29	82.20	133.29	82.20
	47	85	86	95.50-97.00	94.81-95.98	96.09-96.59	95.27-95.66
X	28	120	76	125.44	78.54	116.29	74.27
	56	77	92	73.34	88.45	81.75-84.25	96.30-98.63
Z	32	88	66	94.79	70.55	94.79	70.55
	108	62	130	45.04-45.05	101.07-101.09	46.05-46.80	102.89-104.24

Graficamente, tais resultados podem ser observados através das respectivas marcações nas Figuras VI.2, VI.4, VI.6, VI.8, VI.10, VI.12, VI.14 e VI.16, codificadas da seguinte forma:

Código	Representação
V	Estimativa Visual
L	Estimativa obtida através do Modelo LL
Q	Estimativa obtida através do Modelo LQ

De posse da Tabela VI.2. de resultados, pode-se destacar alguns itens relevantes de análise:

A proximidade numerica obtida entre as estimativas relativas aos modelos ajustados LINEAR-LINEAR e LINEAR-QUADRATICO, que se mantiveram razoavelmente distanciadas apenas em T-28 e T-64, mantendo intermediariamente, nos dois casos, o valor visual quantificado;

- A proximidade numerica obtida entre as estimativas analiticas e os valores visuais quantificados, o que, de certa forma, acaba por "validar" o subjetivismo destes valores. Deve-se observar tambem, que tal comparação difere razoavelmente em T-64 e Z-108, que coincidentemente correspondem aos conjuntos de observações relativos as maiores inclinações nos padrões de esforço fisico dispendido e, portanto, correspondem também aos menores tamanhos amostrais analisados (N=38);

Ao perceber que exatamente em T-64 e Z-108 as estimativas do modelo LINEAR-QUADRATICO estão numericamente bem abaixo dos valores visuais, pode-se tentar tecer uma hipótese de relacionamento entre estes resultados e aqueles obtidos através

das observações simuladas, nos quais, comentava-se sobre a sub-estimação do parâmetro de intercepto e levantava-se a suspeita de que em grupos de observações de pequeno tamanho, a estimativa do intercepto se faz deslocada para a esquerda, presumivelmente devido ao fato de que o sub-modelo quadrático da segunda fase é um melhor ajustante do que o sub-modelo LINEAR da primeira fase;

- O comportamento gráfico "errático" das estimativas  $\hat{SSQR}^{\#}$  na presença de observações desalinhadas ("outliers"), certamente devido ao critério de estimação utilizado e reforçado pela restrição de continuidade no instante do intercepto, dando margem a se questionar a robustez do procedimento.

### VI.3. DISCUSSÕES

Para um trabalho que inicialmente, se propunha a apresentar um procedimento formal analítico de estimação do parâmetro biológico LA, em substituição aos procedimentos visuais subjetivos utilizados até então, este, se desviou um pouco de seus objetivos e, apresentou-se ao final, tal como proposto através do modelo (5.1), composto por diversos componentes não formais.

Comentou-se ao longo da leitura sobre a adoção de sub-modelos lineares compondo o modelo bi-segmentado, devido a facilidade numérica envolvida na sua utilização, quando, para se tornar uma formalização mais realista e mais próxima do problema

abordado, o modelo deveria estar composto por sub-modelos não lineares. Em agravante, fixaram-se os graus dos polinômios dos sub-modelos lineares adotados, desfigurando o carácter formal analítico perseguido, visto que qualquer outra opção utilizada, implicaria na obtenção de resultados discordantes.

Entretanto, tendo em vista o aumento da complexidade analítica implícita na utilização de tais formas funcionais e ainda, observando-se a eficiência (em termos comparativos aos resultados visuais obtidos) das estimativas do Modelo LQ, questiona-se a validade de tal formalização.

Discutiu-se ainda sobre a adoção de um comportamento autorregressivo de ordem 1 para as observações temporais analisadas, quando na verdade tal comportamento se mostrou também, através de testes preliminares, sob a forma autorregressivo de ordem 2. Sobre o procedimento de estimação utilizado para o parâmetro autorregressivo, comentou-se que poderia ter sido "melhorado" em termos de precisão, caso fosse utilizado de forma iterada, porém, argumentou-se que isso faria com que a análise do modelo proposto se tornasse inviável do ponto de vista computacional.

Comenta-se entretanto que alguns dos problemas citados acima certamente poderiam ter sido solucionados caso se adotasse um comportamento estocástico Normal para os resíduos do modelo proposto e se procedesse alguns Testes de Hipótese sobre seus parâmetros.



Finalizando, comenta-se que, apesar de todas as incongruências encontradas ao longo do trabalho, o procedimento de análise do problema biológico abordado mostrou-se sensível, simples e concordante com os resultados obtidos através da análise clássica visual e, portanto, pode ser citado como um procedimento viável na abordagem de situações correlatas. Pode-se também citar em sua defesa, a exigüidade de recursos computacionais exigidos na sua implementação.

Fica claro entretanto, que a pesquisa necessita de continuidade no sentido de que o modelo proposto se adapte melhor ao problema real abordado, e que as "informalidades" encontradas passem a possuir carácter analítico sem que isso implique na perda da viabilidade computacional do modelo proposto. Seguem-se as Figuras mencionadas ao longo deste capítulo.

FIGURA VI.2. Gráfico de dispersão das observações amostrais da resposta VENTILAÇÃO contra o TEMPO, para o indivíduo T em rampa de 28 Watts/min.

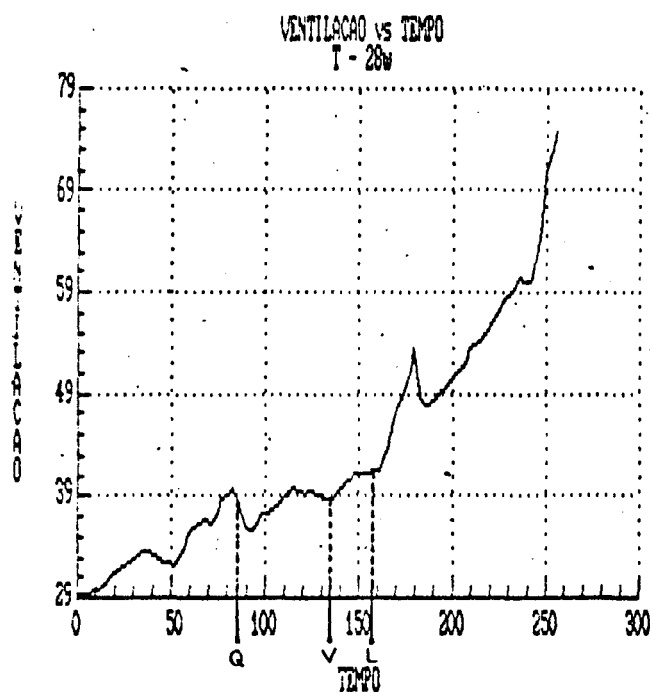


FIGURA VI.3. Gráfico de dispersão dos valores estimados  $\hat{SSQR}^{\#}$  para o indivíduo T em rampa de 28 Watts/min.

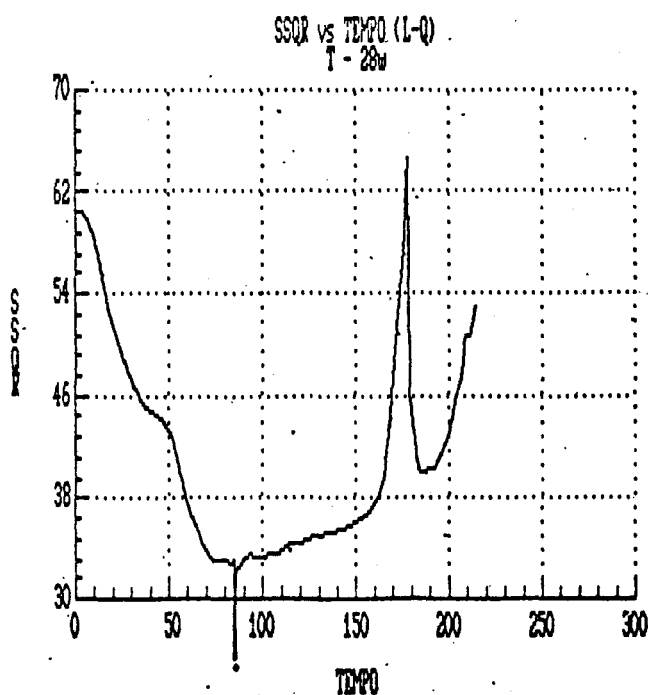


FIGURA VI.4. Gráfico de dispersão das observações amostrais da resposta VENTILAÇÃO contra o TEMPO, para o indivíduo T em rampa de 64 Watts/min.

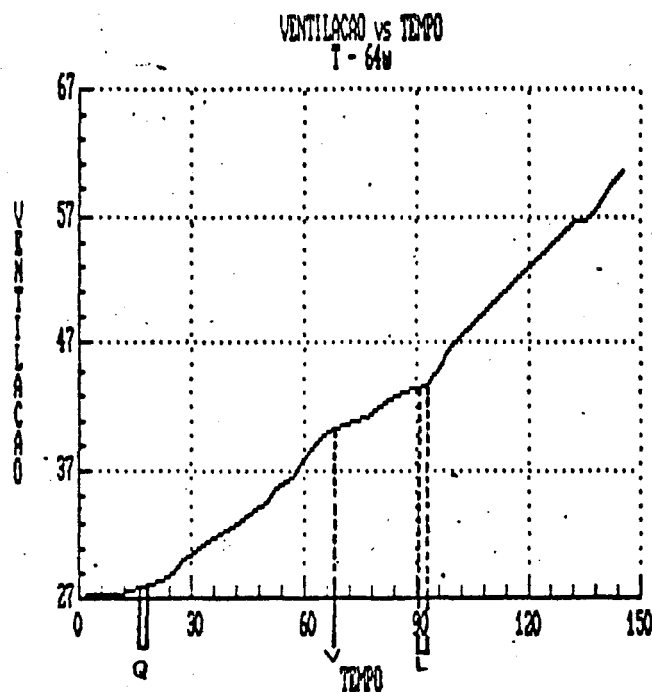


FIGURA VI.5. Gráfico de dispersão dos valores estimados  $\hat{SSQR}^{\#}$  para o indivíduo T em rampa de 64 Watts/min.

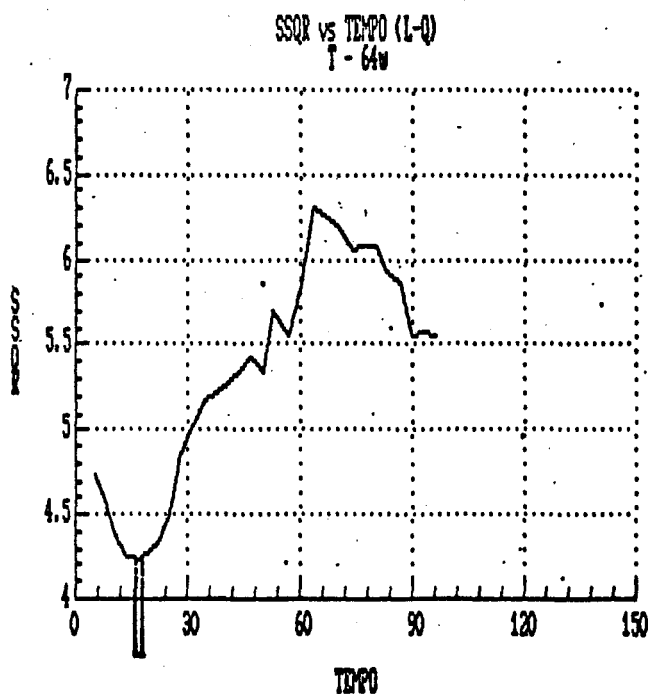


FIGURA VI.6. Gráfico de dispersão das observações amostrais da resposta VENTILAÇÃO contra o TEMPO, para o indivíduo U em rampa de 28 Watts/min.

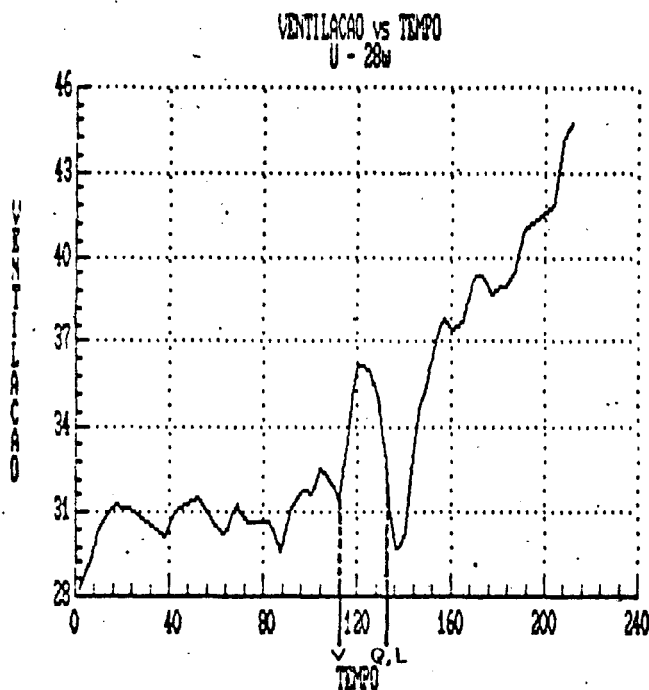


FIGURA VI.7. Gráfico de dispersão dos valores estimados  $\hat{SSQR}^{\#}$  para o indivíduo U em rampa de 28 Watts/min.

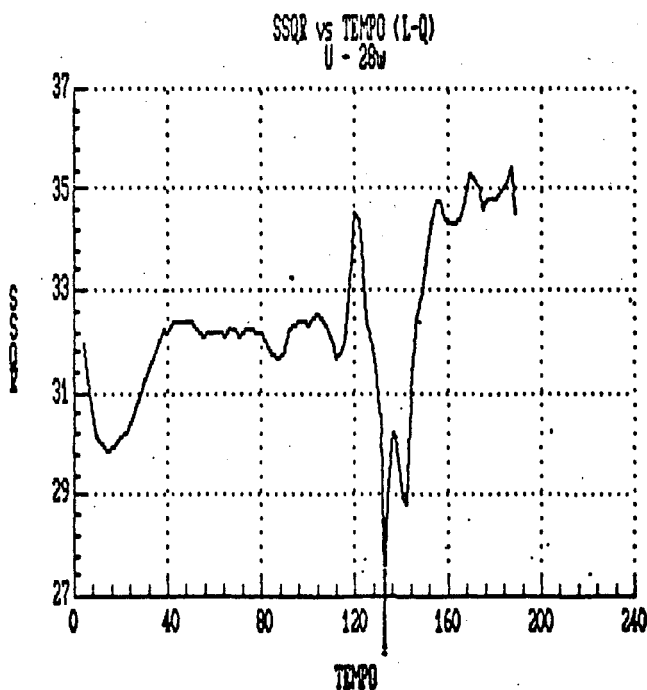


FIGURA VI.8. Gráfico de dispersão das observações amostrais da resposta VENTILAÇÃO contra o TEMPO, para o indivíduo U em rampa de 47 Watts/min.

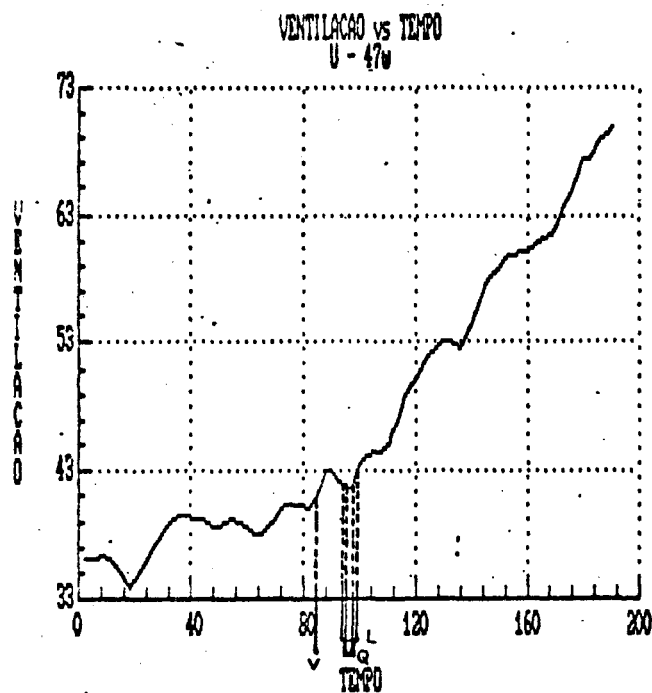


FIGURA VI.9. Gráfico de dispersão dos valores estimados  $\hat{SSQR}^{\#}$  para o indivíduo U em rampa de 47 Watts/min.

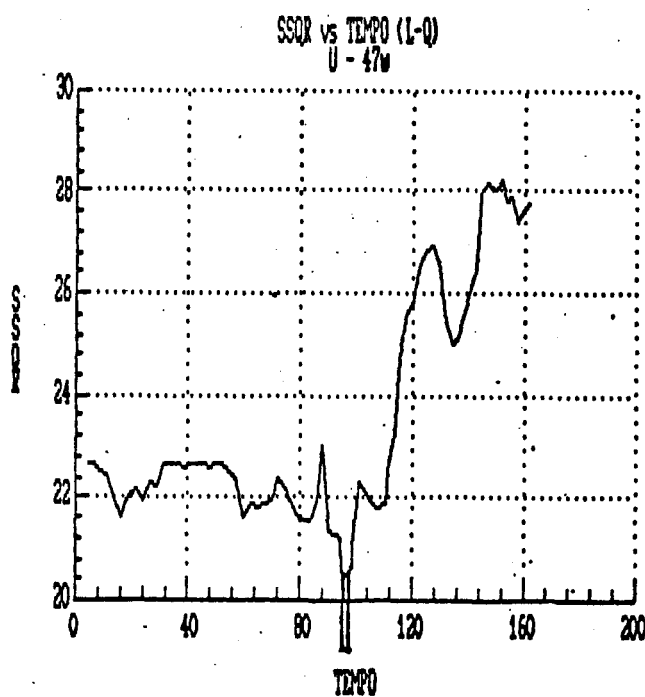


FIGURA VI.10. Gráfico de dispersão das observações amostrais da resposta VENTILAÇÃO contra o TEMPO, para o indivíduo X em rampa de 28 Watts/min.

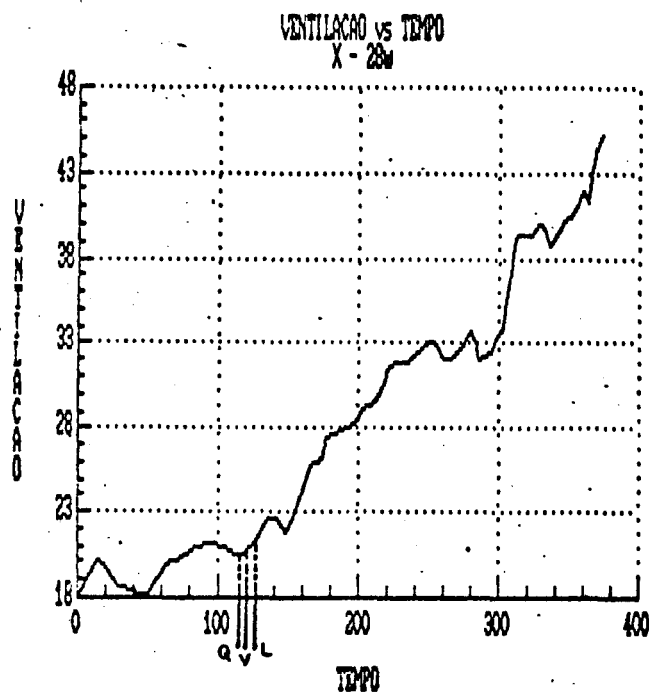


FIGURA VI.11. Gráfico de dispersão dos valores estimados  $\hat{SSQR}^{\#}$  para o indivíduo X em rampa de 28 Watts/min.

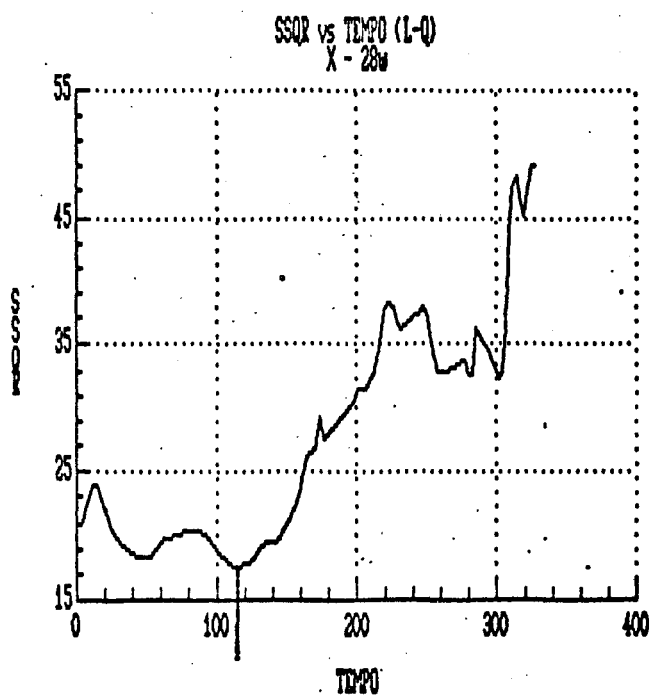


FIGURA VI.12. Gráfico de dispersão das observações amostrais da resposta VENTILAÇÃO contra o TEMPO, para o indivíduo X em rampa de 56 Watts/min.

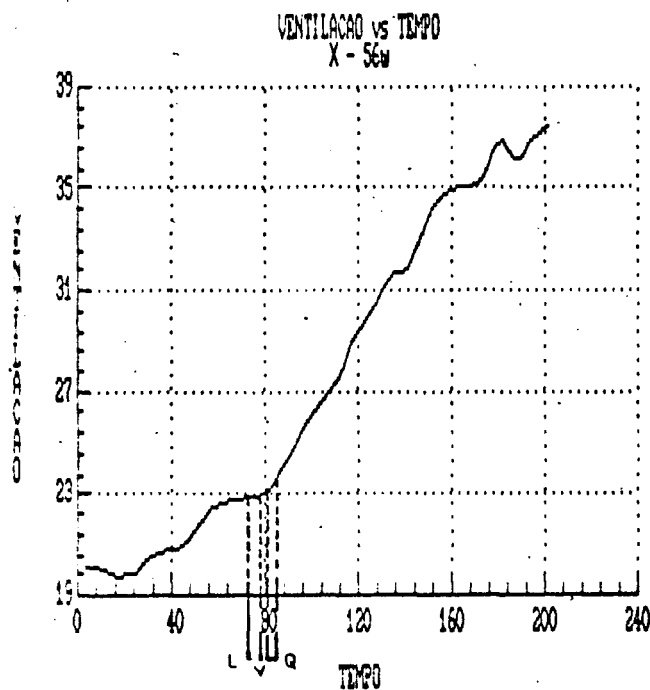


FIGURA VI.13. Gráfico de dispersão dos valores estimados  $\hat{SSQR}^{\#}$  para o indivíduo X em rampa de 56 Watts/min.

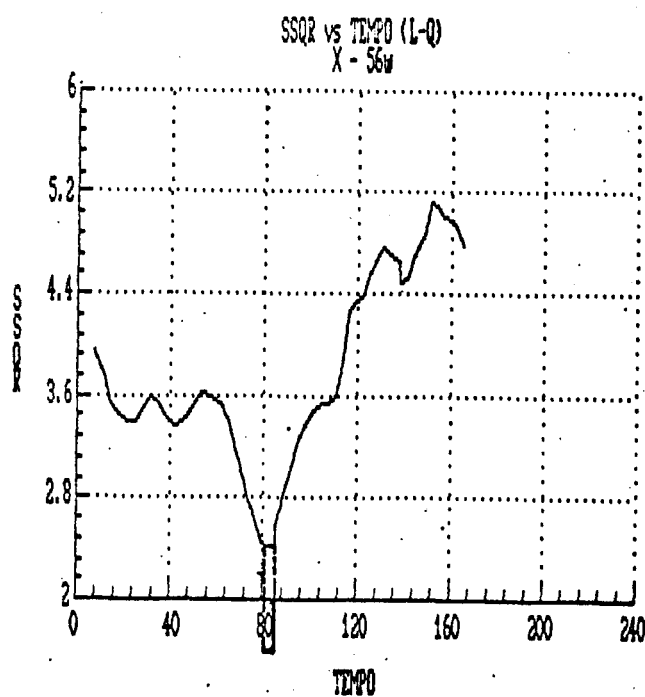


FIGURA VI.14. Gráfico de dispersão das observações amostrais da resposta VENTILAÇÃO contra o TEMPO, para o indivíduo Z em rampa de 32 Watts/min.

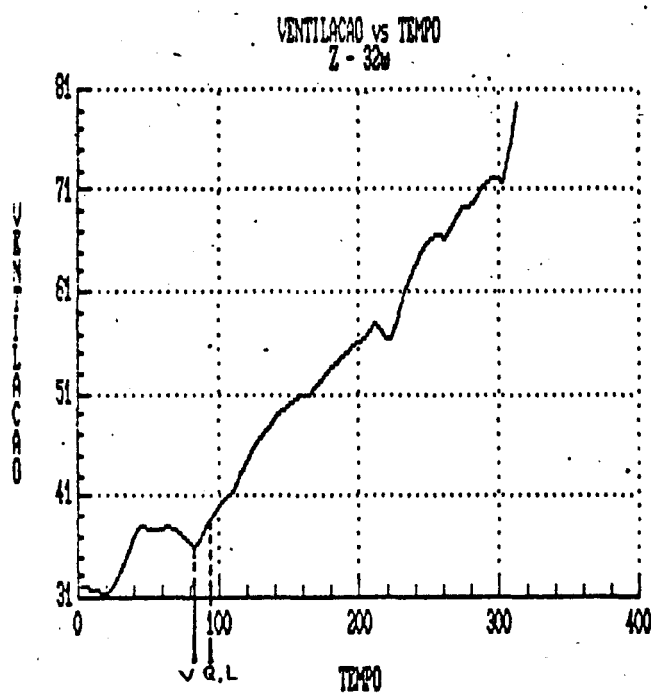


FIGURA VI.15. Gráfico de dispersão dos valores estimados  $\hat{SSQR}^{\#}$  para o indivíduo Z em rampa de 32 Watts/min.

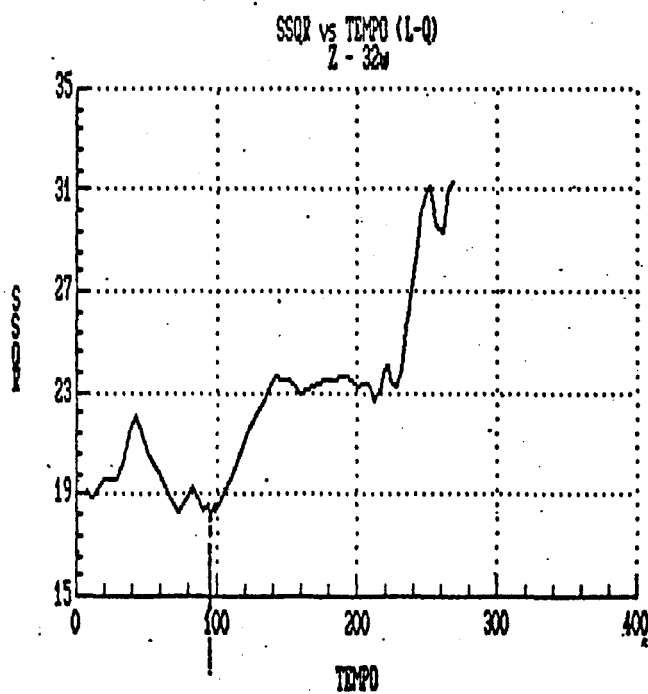




FIGURA VI.16. Gráfico de dispersão das observações amostrais da resposta VENTILAÇÃO contra o TEMPO, para o indivíduo Z em rampa de 108 Watts/min.

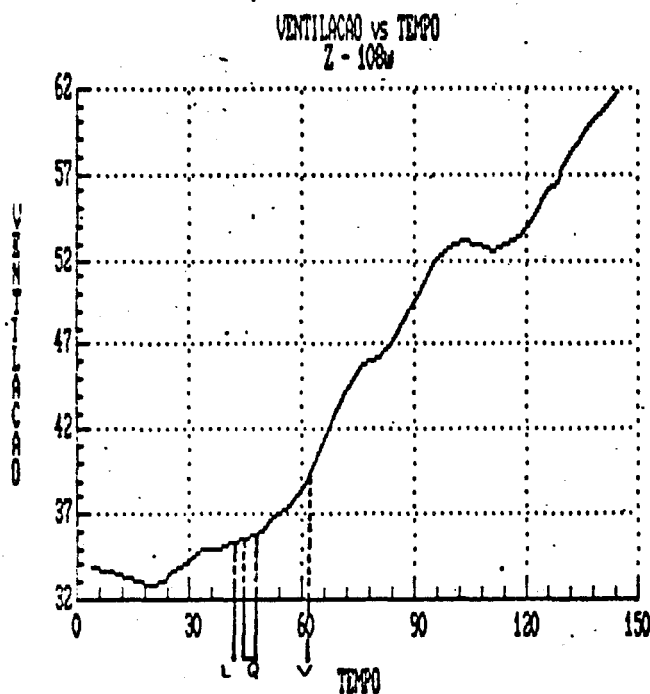
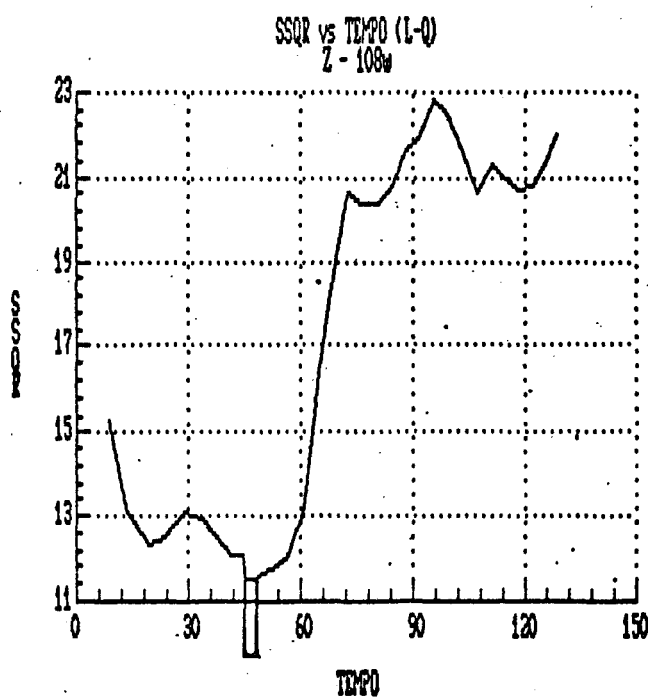


FIGURA VI.17. Gráfico de dispersão dos valores estimados  $\hat{SSQR}^{\#}$  para o indivíduo Z em rampa de 108 Watts/min.



## SUMÁRIO

No presente trabalho, de cunho bioestatístico, procura-se ajustar um modelo de regressão linear bi-segmentado a respostas cardiorespiratórias não invasivas, obtidas de experimentações realizadas em indivíduos sadios submetidos a esforço físico progressivo. O interesse principal do trabalho consiste na estimação do ponto de transição entre os dois sub-modelos lineares considerados, objetivando a quantificação de uma medida fisiológica associada ao estado físico e orgânico individual, denominada Limiar de Anaerobiose (LA).

Para tanto, são revistos alguns métodos estatísticos comumente utilizados no tratamento de modelos de regressão segmentados, dentre os quais, destacam-se os procedimentos iterativos citados por Hudson (1966) e por Lerman (1980). Propõe-se uma versão do procedimento de Lerman (1980) através do critério dos mínimos quadrados generalizados, para o tratamento de medidas sujeitas a erros autocorrelacionados.

O modelo proposto é utilizado, ao final, na análise de algumas observações experimentais conhecidas. Os resultados encontrados são, então, comparados graficamente com resultados prévios provenientes de análises subjetivas clássicas.

## SUMMARY

This work, intended to be applied in biostatistic, looks for a fit of a two phase linear regression model to noninvasive cardiorespiratory responses, obtained from performed on healthy males subjects to incremental physical loads. The main concern of the work consists in the estimation of the joint point between the two respected linear sub-models, intended to measure a physiological gauge related to the physical and organic individual state, called anaerobic threshold (ATD).

In this way, some statistical methodologies, customary available in the treatment of segmented regression models, are reviewed, emphasizing the iterative procedures mentioned by Hudson (1966) and by Lerman (1980). This work suggests a version to the Lerman's (1980) procedure through the generalized least square criterion to analysis of values subjected to autocorrelated errors.

Finally, the suggested model is utilized in the analysis of some experimental observations with known results. The obtained results are grafically compared with previous ones arising in classical subjective studies.

# APÊNDICE 1

INDIVÍDUO T - RAMPA DE 28 W/MIN.

TEMPO (SEG)	VENTILAÇÃO (L/MIN)
-------------	--------------------

1.1499940	29.2500000
3.9500120	29.3600000
7.2500000	29.4600000
9.3999940	29.5600000
11.8500100	29.6100000
15.2000100	30.3400000
18.0900300	31.1600000
21.8400300	31.6700000
24.6400100	32.2300000
27.5000000	32.4500000
30.5900300	32.7600000
34.3400300	33.4400000
37.7500000	33.5300000
40.7899800	33.2500000
43.7999900	32.9400000
46.2000100	32.4800000
49.0000000	32.3800000
51.6500200	32.3000000
54.0499900	32.9200000
56.5000000	33.6900000
58.6500200	34.7700000
62.2000100	35.9900000
65.4000200	36.4200000
68.0000000	36.5200000
71.6500200	36.4800000
74.4400000	37.5400000
76.5900300	38.6800000
80.6900000	39.1400000
82.9400000	39.6600000
85.1500200	38.8400000
87.7000100	37.0400000
89.6900000	36.0300000
91.9000200	35.7700000
93.8400300	35.6000000

cont.

96.0000000	36.3600000
98.2500000	37.3700000
101.4000000	37.5500000
103.5500000	37.7800000
105.9500000	37.9300000
108.3500000	38.4000000
110.7500000	38.9700000
112.9000000	39.5800000
115.0500000	39.8200000
117.2000000	39.5700000
119.1500000	39.4300000
121.0500000	39.1900000
122.9500000	39.3200000
124.9400000	39.3500000
126.9500000	39.1700000
128.7000000	39.0000000
132.1500000	38.7500000
134.4000000	38.6300000
136.4400000	38.9400000
138.3000000	39.3200000
141.7900000	39.8300000
143.9500000	40.4200000
145.7900000	40.6300000
147.0000000	41.2100000
149.5000000	41.2200000
152.8000000	41.2800000
155.4000000	41.3600000
157.2900000	41.4500000
159.0900000	41.5500000
161.0500000	41.5900000
162.8400000	42.5200000
164.9500000	43.5600000
166.5900000	44.7500000
168.5000000	46.3500000
170.4000000	47.6700000
173.9000000	49.0200000
176.0400000	50.4000000
177.8000000	51.4800000
179.0000000	53.3200000

cont.

183.4400000	48.1500000
185.5000000	47.9900000
187.5500000	47.9900000
189.6500000	48.3000000
192.5900000	48.7600000
194.8400000	49.1600000
199.0500000	50.1400000
201.5900000	51.1400000
206.9000000	52.2100000
208.9400000	53.3000000
211.5500000	53.8700000
214.6000000	54.2600000
218.9000000	55.2100000
221.4000000	56.1500000
225.0900000	57.1400000
227.7500000	58.2100000
230.0500000	58.6400000
233.2900000	59.5000000
235.8000000	60.3400000
238.5000000	60.2500000
242.2000000	60.0900000
247.3000000	65.0500000
249.4400000	70.5900000
252.7500000	72.5200000
255.0000000	74.7500000

INDIVÍDUO T - RAMPA DE 64 W/MIN.

TEMPO (SEG)	VENTILAÇÃO (L/MIN)
2.2000120	27.2100000
5.3399960	27.1800000
7.7000120	27.3200000
10.3900100	27.3900000
13.6499900	27.5500000
16.6000100	27.7300000
19.2000100	28.0600000

cont.

22.1499900	28.4300000
24.9500100	29.0800000
27.9899900	29.9800000
31.0900000	30.7500000
34.4500100	31.5800000
37.5400100	32.1200000
40.6400100	32.6300000
43.6900000	33.2600000
46.8400300	33.9300000
50.2500000	34.6300000
52.5499900	35.6500000
56.8400300	36.5900000
59.9500100	37.9200000
63.5399800	39.3500000
66.4000200	40.0400000
70.6500200	40.6900000
73.5900300	41.1100000
76.6900000	41.3900000
80.4000200	42.0600000
82.9400000	42.7300000
86.7899800	43.1300000
89.7000100	43.5500000
92.8400300	43.7600000
96.2999900	45.0900000
98.6400100	46.5500000
132.5900000	56.5600000
135.0900000	56.5100000
137.5400000	57.1800000
139.8400000	58.2600000
142.6900000	59.6400000
145.3400000	60.4100000

INDIVIDUO U - RAMPA DE 28 W/MIN.

TEMPO (SEG)	VENTILACAO (L/MIN)
1.8999940	28.2600000
4.1499940	28.7600000
6.8399960	29.3800000
9.4500120	30.2500000
12.0400100	30.6900000
14.5499900	31.0500000
17.4400000	31.2500000
20.0000000	31.2000000
22.5000000	31.0900000
38.0000000	30.1800000
40.0400100	30.5200000
42.5000000	30.9700000
44.7999900	31.1000000
47.2900100	31.2500000
49.8900100	31.4000000
52.3999900	31.5000000
54.9400000	31.1500000
57.2000100	30.7300000
59.5499900	30.5100000
62.2999900	30.2400000
64.6000100	30.3900000
66.6900000	30.8900000
68.7500000	31.2200000
70.9000200	30.9400000
73.2999900	30.6000000
75.6400100	30.6400000
77.8400300	30.6400000
80.1500200	30.6500000
82.5399800	30.7000000
85.0000000	30.2000000
87.2500000	29.6300000
89.3400300	30.3500000
91.7500000	31.1700000
93.9000200	31.4200000
96.4400000	31.7900000



cont.

98.5900300	31.7700000
100.4000000	31.6400000
102.4000000	32.0800000
104.1900000	32.5400000
106.0500000	32.3600000
108.2500000	32.1800000
110.3400000	31.8400000
112.1900000	31.4000000
113.9000000	32.3800000
115.8400000	33.5300000
117.5900000	34.8300000
119.9000000	36.1200000
122.2500000	36.1800000
124.6900000	36.0900000
128.7500000	35.0300000
130.2900000	33.9000000
132.2500000	32.7600000
133.9400000	30.9100000
136.1900000	29.7500000
137.6500000	29.7200000
140.4500000	30.2800000
142.4000000	31.7700000
144.5000000	33.6300000
146.5900000	34.6700000
148.9000000	35.3500000
150.9400000	36.2000000
153.1000000	36.9100000
155.0000000	37.4000000
156.7500000	37.8700000
158.9000000	37.7300000
160.8000000	37.4800000
162.9000000	37.5500000
165.1500000	37.6700000
167.0900000	38.3900000
169.1900000	39.2400000
171.2000000	39.3400000
173.3000000	39.3300000
175.2000000	39.1100000
177.2500000	38.7300000

cont.

179.1900000	38.8300000
181.0500000	38.9500000
183.0500000	38.9300000
185.3400000	39.2000000
187.1900000	39.4900000
189.0900000	40.1600000
191.2500000	40.9300000
203.7500000	41.8300000
205.5500000	42.6500000
207.5900000	44.1500000
211.6000000	44.7300000

INDIVÍDUO U - RAMPA DE 47 W/MIN.

TEMPO (SEG)	VENTILAÇÃO (L/MIN)
2.3399960	35.9400000
4.7500000	35.9600000
6.7500000	36.1800000
9.0400090	36.4300000
11.1900000	36.0200000
13.4400000	35.5500000
16.0000000	34.7800000
17.9400000	33.8700000
19.9500100	34.3300000
22.0400100	34.9200000
24.1499900	35.7000000
26.6400100	36.7900000
29.1499900	37.6800000
31.3999900	38.6100000
33.8500100	39.1500000
36.5000000	39.5500000
39.0400100	39.4000000
41.5400100	39.1000000
44.0900000	39.0100000
46.5000000	38.8500000
47.9899900	38.6900000

cont.

50.2000100	38.5400000
52.2000100	38.8100000
54.6499900	39.1400000
56.9500100	38.9500000
59.5900000	38.7200000
62.2500000	38.0800000
64.9500100	37.9700000
67.2500000	38.5400000
70.0000000	39.2700000
72.0000000	40.1000000
74.3400000	40.2700000
76.7900100	40.2800000
79.2999900	40.2000000
81.7000100	40.0800000
83.7999900	40.6800000
86.0499900	41.7700000
88.0000000	43.0400000
90.2500000	43.0000000
92.5900300	42.2300000
95.0000000	42.0300000
97.0000000	41.6200000
98.6500200	42.6600000
101.0000000	43.9600000
103.0500000	44.2400000
105.4400000	44.4200000
107.5400000	44.4700000
110.5500000	45.0600000
111.4000000	45.9400000
113.6500000	47.1100000
115.6900000	48.5800000
117.9000000	49.4700000
120.0400000	50.1100000
122.4400000	51.1600000
124.8000000	52.0000000
127.3400000	52.5600000
129.5900000	53.0600000
131.7500000	53.0200000
134.1900000	52.7600000
136.0900000	52.6700000

cont.

138.2500000	53.5400000
140.3000000	54.5500000
142.5000000	55.8200000
144.4000000	57.2800000
146.5900000	58.0500000
149.9500000	58.8200000
151.8500000	59.5300000
153.8000000	59.8100000
155.7500000	59.8000000
157.5000000	59.9700000
159.7500000	60.0600000
161.8000000	60.5100000
164.0500000	61.0300000
166.2500000	61.2500000
168.5900000	61.4800000
170.5000000	62.0200000
172.3400000	63.3300000
175.9000000	64.8000000
178.0500000	66.3000000
179.7500000	67.2600000
182.3000000	67.4100000
184.4400000	68.1700000
186.4400000	68.9900000
188.3000000	69.3800000
190.4400000	69.8100000

INDIVIDUO X - RAMPA DE 28 W/MIN.

TEMPO (SEG)	VENTILAÇÃO (L/MIN)
1.7900090	18.1100000
5.0000000	18.6100000
8.1499940	19.2300000
11.6499900	19.7400000
14.8900100	20.1000000
18.5000000	19.8300000
22.0000000	19.4500000

cont.

25.7000100	19.1300000
29.5400100	18.7100000
33.0900000	18.5500000
36.7500000	18.4200000
40.2000100	18.3200000
43.5000000	18.2600000
47.0400100	18.2300000
50.2900100	18.2200000
53.8999900	18.5600000
56.7900100	18.9500000
60.3999900	19.3800000
63.7500000	19.8600000
67.2900100	20.0200000
70.7000100	20.1400000
74.2000100	20.3500000
77.6400100	20.5000000
80.9500100	20.7000000
84.3400000	20.9100000
87.8500100	21.0100000
91.2900100	21.1100000
94.8999900	21.1100000
98.2000100	21.0800000
101.5400000	21.0300000
105.0400000	20.8600000
108.3400000	20.7200000
112.2000000	20.5900000
116.2900000	20.5300000
119.4400000	20.6000000
122.4500000	20.8300000
125.9500000	21.1000000
129.2900000	21.5500000
132.4400000	22.0900000
135.8500000	22.5300000
139.0000000	22.6700000
142.2900000	22.5200000
145.5900000	22.2400000
148.3900000	21.8500000
151.1000000	22.1600000
153.3500000	22.5400000

cont.

156.2000000	23.1600000
158.7900000	24.1500000
161.7000000	24.8000000
164.8500000	25.4900000
167.8900000	25.9100000
170.7500000	26.0300000
173.9000000	26.1300000
176.9500000	27.5000000
197.4500000	28.3400000
200.4000000	28.7400000
203.4400000	29.1400000
206.2500000	29.2200000
209.0000000	29.2800000
211.6400000	29.6000000
214.5400000	29.9200000
218.0400000	30.5500000
220.4000000	31.3300000
223.0500000	31.6500000
226.0900000	31.8000000
228.9400000	31.8900000
231.7500000	31.7900000
234.4000000	31.8900000
236.9000000	32.1400000
239.4000000	32.3200000
242.0500000	32.5400000
244.7900000	32.7300000
247.5000000	32.9200000
250.1500000	33.1900000
252.6900000	33.1200000
255.3000000	32.8100000
257.8500000	32.4100000
260.1900000	31.9500000
262.5500000	31.9700000
265.1900000	32.0400000
268.0000000	32.2000000
270.1400000	32.4400000
272.6900000	32.6700000
275.2000000	33.0700000
277.9400000	33.4900000

cont.

280.1900000	33.8100000
282.7000000	33.1000000
285.2500000	32.1100000
294.4400000	32.4700000
296.8000000	33.0000000
299.2500000	33.4100000
301.6500000	33.7200000
304.5000000	35.0100000
306.8400000	36.4200000
309.0000000	37.7500000
311.1500000	39.1400000
314.2500000	39.3000000
317.0900000	39.2900000
319.9400000	39.3500000
322.7500000	39.3100000
324.9400000	39.6800000
327.1500000	40.1000000
329.2000000	39.9500000
331.4500000	39.7700000
333.6500000	39.3500000
335.7500000	38.8100000
337.8400000	39.0500000
339.9000000	39.3600000
342.1500000	39.5700000
344.5000000	39.9600000
346.5400000	40.1700000
348.7000000	40.3400000
350.7500000	40.5100000
353.2000000	40.8000000
355.3000000	41.1100000
357.3400000	41.4600000
359.3400000	41.8300000
361.5000000	41.6300000
363.4500000	41.3600000
365.6500000	42.6900000
367.6900000	44.1600000
369.6500000	44.6400000
372.9500000	45.2300000

INDIVÍDUO X - RAMPA DE 56 W/MIN.

TEMPO (SEG)	VENTILAÇÃO (L/MIN)
3.7500000	20.0800000
7.5899960	20.0900000
11.5400100	20.0100000
14.4500100	19.8900000
18.1900000	19.7400000
21.5900000	19.7900000
25.0400100	19.9000000
28.2399900	20.1900000
31.7999900	20.5500000
35.1400100	20.7100000
38.9500100	20.8700000
42.8999900	20.8800000
46.2900100	21.1600000
49.6499900	21.5500000
53.3900100	21.9500000
57.5900000	22.4300000
62.2500000	22.6000000
65.4500100	22.7400000
69.1499900	22.8300000
72.8400000	22.8700000
76.2000100	22.9000000
79.2500000	23.0400000
82.2999900	23.1800000
85.5000000	23.7000000
88.4500100	24.2900000
91.5900000	24.7700000
94.7000100	25.3000000
98.2999900	25.8000000
101.5900000	26.2900000
105.0900000	26.6700000
107.5000000	27.0300000
110.4500000	27.3500000
113.2000000	27.9400000
116.3000000	28.7200000
119.1500000	29.2700000



cont.

122.0000000	29.5500000
124.6900000	30.0800000
127.9000000	30.5800000
130.9000000	31.1100000
135.1900000	31.6700000
137.6500000	31.7200000
138.0900000	31.7100000
141.1500000	31.8700000
143.6500000	32.4100000
146.2900000	32.9500000
148.8400000	33.5500000
151.4400000	34.2400000
154.1500000	34.5100000
156.8000000	34.7500000
160.5000000	34.9300000
162.8400000	35.0300000
165.2500000	35.0800000
168.0900000	35.1000000
170.7000000	35.2100000
173.0900000	35.3200000
175.5000000	35.8900000
177.6500000	36.5300000
179.9000000	36.7900000
182.1900000	36.8200000
184.8000000	36.4300000
187.1900000	36.1400000
189.4500000	36.2300000
191.5500000	36.4200000
193.6500000	36.8200000
196.2000000	37.0700000
199.0900000	37.2600000
201.2000000	37.4500000

INDIVÍDUO Z - RAMPA DE 32 W/MIN.

TEMPO (SEG)	VENTILAÇÃO (L/MIN)
2.8400270	31.9600000
7.5900270	31.9400000
8.0900270	31.8100000
12.7500000	31.5700000
19.8400300	31.2000000
24.2000100	31.5900000
29.5900300	32.9900000
33.6900000	34.6100000
38.0399800	36.2500000
41.9000200	37.5100000
46.8900100	37.8600000
51.5399800	37.7100000
56.0399800	37.5700000
60.0399800	37.7100000
64.5499900	37.8300000
69.0900300	37.6100000
72.7999900	37.3700000
78.2899800	36.6300000
82.5900300	35.9100000
87.0000000	36.6500000
90.5499900	37.5300000
94.7899800	38.6400000
98.5900300	39.8600000
102.5500000	40.4800000
107.0500000	41.0700000
111.0500000	41.5800000
114.7500000	42.8600000
118.8000000	44.0600000
122.2500000	45.0200000
126.0400000	46.0500000
129.8000000	46.6900000
133.8000000	47.2600000
138.3400000	48.1000000
142.3000000	49.0900000
146.8400000	49.6000000

cont.

151.6500000	50.1100000
155.8400000	50.5700000
159.5000000	50.7800000
164.2500000	50.9900000
168.5000000	51.7100000
172.8400000	52.4600000
177.0000000	53.0800000
180.6000000	53.7200000
184.5400000	54.1900000
188.1500000	54.6100000
191.6500000	55.2300000
195.8000000	55.8400000
200.3000000	56.1300000
204.3400000	56.4000000
208.2500000	57.1700000
212.1500000	57.9700000
216.0900000	57.2400000
219.7500000	56.3700000
222.7500000	56.4200000
223.4000000	56.3600000
227.7500000	58.3500000
231.5500000	60.7500000
235.0400000	61.7800000
238.3000000	63.2900000
242.1900000	64.3100000
245.5000000	65.1300000
249.1400000	65.9300000
251.8000000	66.3000000
255.3000000	66.5900000
258.3400000	66.4700000
261.7900000	66.2300000
264.9400000	67.1200000
268.2000000	68.1000000
271.3500000	68.6500000
274.5900000	69.3100000
278.0500000	69.5300000
281.5500000	69.6300000
285.3000000	70.5800000
288.6400000	71.5600000

cont.

291.9400000	71.8500000
295.3000000	72.1700000
298.6900000	72.1200000
302.4400000	71.9100000
306.1000000	74.0600000
309.6500000	76.4700000
313.0500000	79.5100000

INDIVIDUO Z - RAMPA DE 108 W/MIN.

TEMPO (SEG)	VENTILAÇÃO (L/MIN)
4.0000000	33.8600000
8.7500000	33.7400000
13.5900000	33.2900000
19.4500100	32.7700000
23.7500000	33.1000000
24.2999900	33.4900000
28.9500100	34.1400000
33.5400100	34.8300000
37.5400100	34.9400000
41.2999900	35.3000000
45.0400100	35.6200000
49.4500100	36.0500000
52.6000100	36.8600000
56.9500100	37.4600000
60.7500000	38.7100000
64.8999900	40.8800000
68.9500100	42.8900000
72.7000100	44.6700000
76.2000100	45.9000000
80.7500000	46.1900000
84.6400100	47.3200000
88.1499900	48.7900000
91.8500100	50.3000000
95.3900100	51.9800000
99.2500000	52.6500000

cont.

103.3400000	53.2000000
107.2500000	53.0300000
111.3400000	52.5900000
114.6500000	53.0200000
118.4400000	53.4500000
122.4000000	54.5400000
125.9400000	56.0200000
128.8400000	56.6200000
129.3000000	57.2400000
133.4500000	58.6700000
137.0400000	59.9400000
141.5000000	60.8500000
144.4000000	61.8100000

## APÊNDICE 2.

### SOLUÇÃO DO MODELO DE REGRESSÃO LINEAR COM RESTRIÇÕES NOS PARÂMETROS

Considere o modelo de regressão:

$$Y = X\beta + \varepsilon$$

onde,  $X_{N \times p}$  é uma matriz de posto completo ( $p$ );

$\beta_{p \times 1}$  é um vetor de parâmetros;

$\varepsilon_{N \times 1}$  é um vetor de valores aleatórios, tal que:

$$E(\varepsilon) = 0 \quad \text{e} \quad V(\varepsilon) = E(\varepsilon \varepsilon') = I\sigma^2$$

Seja  $R\beta = r$  o conjunto de restrições impostas aos parâmetros, onde  $R_{m \times p}$  é a matriz das relações lineares entre os parâmetros e  $r$  é um vetor  $m$ -dimensional de constantes. Se as  $m$  relações lineares são independentes, então, o posto de  $R$  é igual a  $m$ , tal que,  $m \leq p \leq N$ .

Para obter o estimador  $\hat{\beta}^+$  de  $\beta$ , devemos minimizar:

$$SQR^+ = \varepsilon' \varepsilon = (Y - X\beta)'(Y - X\beta)$$

condicionado à restrição  $R\beta = r$ .

Utilizando o método dos multiplicadores de Lagrange, forma-se a função:

$$S = (Y - X\beta)'(Y - X\beta) + 2\lambda'(R\beta - r)$$

onde,  $\lambda_{m \times 1}$  é o vetor dos multiplicadores de Lagrange.

Desenvolvendo  $S$  e resolvendo o seguinte sistema:

$$\begin{cases} \partial S / \partial \beta = 0 \\ \partial S / \partial \lambda = 0 \end{cases}$$

Após as simplificações necessárias, obtém-se:

$$\begin{cases} -X'Y + (X'X)\hat{\beta}^+ + R'\hat{\lambda}^+ = 0 & (7.1) \\ R\hat{\beta}^+ - r = 0 & (7.2) \end{cases}$$

onde  $\hat{\beta}^+$  e  $\hat{\lambda}^+$  são soluções das equações acima.

Define-se o sistema formado pelas equações (7.1) e (7.2) como o Sistema de Equações Normais. Em notação matricial tem-se:

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta}^+ \\ \hat{\lambda}^+ \end{bmatrix} = \begin{bmatrix} X'Y \\ r \end{bmatrix}$$

No caso em que  $(X'X)$  é não singular, pré-multiplica-se a equação (7.1) por  $R(X'X)^{-1}$ , obtendo-se:

$$\begin{aligned} -R(X'X)^{-1}X'Y + R(X'X)^{-1}(X'X)\hat{\beta}^+ + R(X'X)^{-1}R'\hat{\lambda}^+ &= 0 \quad \text{ou melhor,} \\ -R(X'X)^{-1}X'Y + R\hat{\beta}^+ + R(X'X)^{-1}R'\hat{\lambda}^+ &= 0 \end{aligned}$$

Como  $(X'X)^{-1}X'Y = \hat{\beta}$ , no modelo sem restrição, obtém-se:

$$R\hat{\beta} + R\hat{\beta}^+ + R(X'X)^{-1}R'\hat{\lambda}^+ = 0$$

Substituindo esse resultado em (7.2) e, notando que  $RC(X'X)^{-1}R'$  é não singular, (como pode-se demonstrar), vem:

$$RC(X'X)^{-1}R'\hat{\lambda}^+ = R\hat{\beta} - r \quad \text{e, portanto,}$$

$$\hat{\lambda}^+ = [RC(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)$$

que, substituindo em (7.1) obtém-se:

$$-X'Y + X'X\hat{\beta}^+ + R'[RC(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) = 0$$

$$X'X\hat{\beta}^+ = X'Y - R'[RC(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)$$

pré-multiplicando por  $(X'X)^{-1}$ , tem-se:

$$\hat{\beta}^+ = \hat{\beta} - (X'X)^{-1}R'[RC(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \quad (7.3)$$

Que é o estimador de mínimos quadrados restrito, igual ao estimador de mínimos quadrados irrestrito  $\hat{\beta}$ , mais uma combinação linear das diferenças  $(R\hat{\beta} - r)$ .

A estimativa da soma de quadrados de resíduos para o modelo com restrição, é dada por:

$$\begin{aligned} \hat{SQR}^+ &= (Y - X\hat{\beta}^+)'(Y - X\hat{\beta}^+) \\ &= Y'Y - Y'X\hat{\beta}^+ - \hat{\beta}^{+'}X'Y + \hat{\beta}^{+'}X'X\hat{\beta}^+ \\ &= Y'Y - 2Y'X\hat{\beta}^+ + \hat{\beta}^{+'}X'X\hat{\beta}^+ \end{aligned}$$

Substituindo  $\hat{\beta}^+$  por (7.3), tem-se:

$$\begin{aligned} \hat{SQR}^+ &= Y'Y - 2Y'X \left\{ \hat{\beta} - (X'X)^{-1}R'[RC(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \right\} + \\ &\quad + \left\{ \hat{\beta} - (X'X)^{-1}R'[RC(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \right\}'(X'X) \cdot \\ &\quad \cdot \left\{ \hat{\beta} - (X'X)^{-1}R'[RC(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \right\} \end{aligned}$$



$$\begin{aligned}
&= Y'Y - 2Y'X\hat{\beta} + 2Y'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) + \\
&\quad + \hat{\beta}'(X'X)\hat{\beta} - \hat{\beta}'(X'X)(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) - \\
&\quad - (R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}(X'X)\hat{\beta} + \\
&\quad + (R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}(X'X)(X'X)^{-1} \\
&\quad \cdot R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)
\end{aligned}$$

Lembrando que,  $\hat{\beta}' = Y'X(X'X)^{-1}$  e depois de algumas simplificações:

$$\begin{aligned}
\hat{SQR}^* &= Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'(X'X)\hat{\beta} + 2\hat{\beta}'R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) - \\
&\quad - 2\hat{\beta}'R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) + \\
&\quad + (R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \\
&= Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'(X'X)\hat{\beta} + (R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)
\end{aligned}$$

e finalmente,

$$\hat{SQR}^* = \hat{SQR} + (R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)$$

onde,

$$\begin{aligned}
\hat{SQR} &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\
&= Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'(X'X)\hat{\beta} \\
&= Y'Y - Y'X\hat{\beta}
\end{aligned}$$

que é a estimativa da soma de quadrados de resíduos do modelo irrestrito.

## APÊNDICE 3

### ESTUDO COMPARATIVO DE RESULTADOS

É comum encontrar-se na literatura bio-médica específica, dedicada ao assunto em voga, estudos comparativos de resultados, sempre que uma nova metodologia de quantificação do LA é proposta.

No presente caso, comenta-se que tais estudos seriam questionáveis, visto não existir qualquer parâmetro de comparação entre um resultado obtido visual e subjetivamente, levando em consideração experiências prévias, e outro, obtido a partir da formalização estatística de um determinado critério fisiológico pré-estabelecido.

Questiona-se portanto se a proximidade numérica destes dois resultados validaria a estimativa estatística ou a estimativa visual quantificada pelo pesquisador, ou então, se simplesmente não faria qualquer sentido.

Apesar desta controvérsia, tal estudo foi realizado sobre o conjunto de resultados estimados através do modelo proposto e outros resultados, obtidos visualmente por pesquisadores interessados. O critério de comparação adotado foi o cálculo do Coeficiente de Correlação Linear de Pearson ( $r$ ) entre tais resultados, que é dado por:

$$r = \frac{n.(\Sigma xy) - (\Sigma x)(\Sigma y)}{(\sqrt{n.(\Sigma x^2) - (\Sigma x)^2}).(\sqrt{n.(\Sigma y^2) - (\Sigma y)^2})}$$

Em uma primeira etapa calculou-se  $r$  entre o conjunto de resultados provenientes do Modelo LQ e aqueles resultados obtidos por Martins (1986), que foram apresentados na Tabela VI.2. O valor calculado,  $r = 0.36$ , evidenciou um fraco relacionamento linear positivo entre tais resultados, pondo em dúvida o grau de coincidência entre as estimativas visuais subjetivas e as estimativas provenientes do modelo estatístico proposto.

Tendo em vista que as estimativas visuais quantificadas por Martins (1986) possuíam um certo vício metodológico intencional, decorrente dos objetivos pretendidos em seu trabalho, obteve-se quatro novos conjuntos de valores visuais, quantificados a partir dos gráficos das mesmas medidas utilizadas ao longo deste trabalho (Figuras VI.2, VI.4, VI.6, VI.8, VI.10, VI.12, VI.14 e VI.16), para os quais, repetiu-se o cálculo do coeficiente de correlação linear.

Os dois primeiros conjuntos foram provenientes de valores quantificados por médicos experientes, que foram instruídos a identificar o valor do LA nos gráficos, levando em consideração apenas os critérios fisiológicos citados na literatura. Para tais conjuntos, calculou-se  $r = 0.32$  e  $r = 0.41$ , novamente evidenciando um grau fraco de relacionamento linear entre os resultados visuais e os resultados estatísticos.

Os dois conjuntos seguintes foram quantificados por matemáticos numéricos que foram instruídos a identificar em cada gráfico, um ponto de interseção contínuo entre uma reta e uma curva polinomial de grau desconhecido entre 1 e 2. Para estes conjuntos, calculou-se  $r = 0.76$  e  $r = 0.84$  e, desta vez, obteve-se um grau forte de relacionamento linear entre os resultados comparados.

Comenta-se que tal discriminação constou do delineamento do estudo como uma tentativa de levantar suspeitas sobre possíveis diferenças nos critérios de quantificação visual utilizados. Pressupunha-se que, mesmo instruídos a fazerem uso apenas dos critérios citados, os médicos dispunham de certas informações biológicas adicionais relevante para a quantificação do LA, enquanto que os matemáticos, deveriam se fixar apenas no critério descrito e, com isso, indicariam resultados bem mais próximos daqueles obtidos através do modelo estatístico.

Apesar de o tamanho amostral não ser suficiente para a tomada de qualquer decisão estatística, os cinco valores de  $r$  calculados, parecem reforçar tal hipótese.

De qualquer forma, a estimação do LA através de um modelo estatístico, pretende a formalização de toda informação adicional relevante, o que implica que qualquer diferença razoavelmente grande (levando em consideração o subjetivismo implícito no valor visual) entre o resultado obtido através deste modelo e o

resultado visual quantificado pelo médico experiente, significa a necessidade da melhor especificação do modelo. Tudo isso, alimenta a continuidade da pesquisa no sentido de se propôr um modelo que melhor se adeque ao problema real em questão.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BACON, D.W. ; WATTS, D.G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, 58(3), 525-34.
- BECKMAN, R.J. ; COOK, R.D. (1979). Testing for two-phase regression. *Technometrics*, 21(1), 65-9.
- CAIOZZO, V.J. et al. (1982). A comparison of gas exchange indices used to detect the anaerobic threshold. *J. Appl. Physiol.*, 53(5), 1184-189.
- DACHS, J.N.W. ; CARVALHO, J.F. (1984). Diagnóstico em regressão. Notas de mini curso. 6º SINAPE, UFRJ, Rio de Janeiro (RJ), pp.86.
- DAVIS, J.A. et al. (1979). Anaerobic threshold alternations caused by endurance training in middle-age men. *J. Appl. Physiol.*, 46, 1039-46.
- DEAVER, W.L. et al. (1985). Improved detection of lactate threshold during exercise using a log-log transformation. *J. Appl. Physiol.*, 59(6), 1936-940.
- ESTERBY, S.R. ; EL-SHAARAWI, A.H. (1981). Inference about the point of change in a regression modelo. *Appl. Statist.*, 30(3), 277-85.
- FEDER, P.I. (1975a). On asymptotic distribution theory in segmented regression problems - identified case. *Ann. Statist.*, 3, 49-83.
- \_\_\_\_\_ (1975b). The log likelihood ratio in segmented regression. *Ann. Statist.*, 3, 84-87.
- FERREIRA, P.E. (1975). A Bayesian analysis of a switching regression model: Know number of regimes. *JASA*, 70(350), 370-74.

- GALLANT, A.R. ; FULLER, W.A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. JASA, 68(341), 144-47.
- GLADDEN, L.B. et al. (1985). Gas exchange and lactate anaerobic thresholds: inter and intraevaluator agreement. J. Appl. Physiol., 58(6), 2082-89.
- GREEN, H.J. et al. (1983). Anaerobic threshold, blood lactate, and muscle metabolites in progressive exercises. J. Appl. Physiol., 54(4), 1032-38.
- HAWKINS, D.N. (1976). Point estimation of the parameters of piecewise regression models. Appl. Statist., 25, 51-57.
- HINKLEY, D.V. (1969). Inference about the intersection in two-phase regression. Biometrika, 56(3), 495-504.
- \_\_\_\_\_ (1971). Inference in two-phase regression. JASA, 66(336), 736,43.
- HOFFMAN, R. ; VIEIRA, S. (1972). Análise de regressão. Uma introdução à econometria. Ed. Hucitec, Ed. USP - São Paulo (SP), pp. 339.
- HUDSON, D.J. (1966). Fitting segmented curves whose join points have to be estimated. JASA, 61, 1097-129.
- JOHNSTON, J. (1971). Métodos econométricos. Ed. Atlas, São Paulo (SP), pp. 318.
- LEITE, C.R.C. (1982). Ajuste de dados por modelos segmentados. Dissertação de Mestrado. ICMSC - USP, São Carlos (SP), pp. 76.
- LERMAN, P.M. (1980). Fitting segmented regression models by grid-search. Appl. Statist., 29(1), 77-84.

- MARTINS, L.E.B. (1986). Utilização de análise de sistemas no estudo da resposta cardiorespiratória ao esforço físico dinâmico. Dissertação de Mestrado. FMRP-USP, Ribeirão Preto(SP), pp.147.
- ORR, G.W. et al. (1982). A computer linear regression model to determine ventilatory anaerobic threshold. J. Appl. Physiol., 52(5), 1349-352.
- QUANDT, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. JASA, 53, 873-80.
- ROBINSON, D.E. (1964). Estimates for the points of intersection of two polynomial regressions. JASA, 59, 214-24.
- SALAZAR, D. et al. (1981). Parameter changes in a regression model with autocorrelated errors. Commun. Statist. - Theor. meth., A10(17), 1751-58.
- SANCHES, D. (1986). Análise dos efeitos do treinamento anaeróbico sobre o efluxo e remoção do ácido láctico e mudanças na curva do limiar aeróbico-anaeróbico. Dissertação de Mestrado. EEFD-UFRJ, Rio de Janeiro (RJ), pp.65.
- SEBER, G.A.F. (1977). Linear Regression Analysis. John Wiley, Inc. New York, pp.465.
- SHABAN, S.A. (1980). Change point problem and two-phase regression: an annotated bibliography. Intern. Statist. Rev. 48, 83-93.
- SMITH, P.L. (1979). Splines as a useful and convenient statistical tool. Amer. Statist., 26(3), 59-61.



- SPRENT, P. (1961). Some hypotheses concerning two-phase regression lines. *Biometrics*, 17, 634-45.
- THEIL, H. (1971). Principles of econometrics. John Wiley, Inc. New York, pp.736.
- VINOD, H.D. ; ULLAH, A. (1981). Recent advances in regression methods. Marcel Dekker, Inc. New York, pp.361.
- WASSERMAN, K. ; McILROY, M.B. (1964). Detecting the threshold of anaerobic metabolism. *Amer. Jour. of Cardiol.*, 14, 844-52.
- WASSERMAN, K. et al. (1967). Interaction of physiological mechanism during exercise. *J. Appl. Physiol.*, 22(1), 71-85.
- WASSERMAN, K. et al. (1973). Anaerobic threshold and respiratory gas exchange during exercise. *J. Appl. Physiol.*, 35(2), 236-43.
- WASSERMAN, K. (1986). Anaerobiosis, lactate and gas exchange during exercise: the issues. *Federation Proc.*, 45, 2904-909.
- WHIPP, B.J. et al. (1981). A test to determine parameters of aerobic function during exercise. *J. Appl. Physiol.*, 50(1), 217-21.
- WONNACOTT, R.J. ; WONNACOTT, T.H. (1970). Econometrics. John Wiley, Inc. New York, pp.445.
- YEH, M.P. et al. (1983). Anaerobic threshold: problems of determination and validation. *J. Appl. Physiol.*, 55(4), 1178-186.